An adjusted payoff-based procedure for normal form games

Mario Bravo

Université Pierre et Marie Curie – Paris 6
Institut de Mathématiques de Jussieu
Combinatoire et Optimisation
4 place Jussieu, 75252 Paris cedex 05.
bravo@math.jussieu.fr

Abstract

We study a simple adaptive model in the framework of an N-player normal form game. The model consists of a repeated game where the players only know their own strategy space and their own payoff scored at each stage. The information about the other agents (actions and payoffs) is unknown. In particular, we consider a variation of the procedure studied by Cominetti et al. [9] where, in our case, each player is allowed to use the number of times she has played each action to update her strategy. The resultant stochastic process is analyzed via the so-called ODE method from stochastic approximation theory. We are interested in the convergence of the process to rest points of a related continuous dynamics. Results concerning almost sure convergence and convergence with positive probability are obtained and applied to a traffic game. Also, we provide some examples where convergence occurs with probability zero. Finally, we verify that part of the analysis holds when players are facing a random environment.

Keywords: Normal form games, Learning, Adaptive dynamics, Stochastic approximation
MSC2010 Subject classification: Primary: 91A26, 91A10 ; Secondary: 62L20, 93E35

1 Introduction

This paper studies an adaptive model for an N-player repeated game. We consider boundedly rational players that adapt using simple behavioral rules based on past experiences.

The decision that a player can make at each stage hinges on the amount of information available. Several approaches have been proposed depending on the information that agents can gather over time. Fictitious play (see Brown [7], Fudenberg and Levine [14]) is one of the best studied procedures. Players adapt their behavior by performing best responses to the opponent's average past play over time. In this case, each player needs to know her own payoff function and to receive complete information about the other players' moves. A less restrictive framework is when each player is informed of all the possible payoffs that she could have received by using alternative moves. The exponential procedure (Freund and Shapire [13]) is, among others, one example of this kind of adaptive processes.

We are interested in a less informative context. Players make no anticipation on the opponents' behavior and we assume that they have no information on the structure of the game. This means that agents have only their own space of actions and past realized payoffs to react to the environment. We suppose that players are given a rule of behavior (a *decision rule*) which depends on a state variable. The state variable is updated by a possibly time-dependent rule (an *updating rule*) based on the history of play and current observations.

A widely studied model in this framework is the *cumulative reinforcement learning* procedure where players conserve a vector perception (the state variable), in which each coordinate of the vector represents how a move performs. The updating rule is defined by adding the payoff received to the component of the previous vector perception corresponding to the move actually played, and keeping the other components unaltered for the unused moves. The decision rule is given by the normalization of this perception vector assuming that payoffs are positive. Several results for the convergence (and nonconvergence) of players' mixed actions have been obtained (see Beggs [1], Börgers and Sarin [5], Laslier et al. [18], and a normalized version by Posch [22] for the 2-player game framework and Erev and Roth [12] for experimental results). In Cominetti et al. [9] the authors study a model in the same spirit mainly in the case of Logit rule decision (which allows nonpositive payoffs) in the N-player case. Players update the perception vector by performing an average between the new payoff received and the former perception. Conditions are given to ensure the convergence to a Nash equilibrium of a perturbed version of the game. A similar model is studied by Leslie and Collins [19], where results concerning 2-player games are obtained. Another approach in this information framework is developed by Hart and Mas-Colell [16] where the analysis focuses on the convergence of the empirical frequency of play instead of the long-term behavior of the mixed strategy. Using techniques based on consistent procedures (see Hart and Mas-Colell [15]), it is shown that, for all games, the set of correlated equilibria is attained.

We consider here a particular updating rule where players keep a perception vector that is updated, on the coordinate corresponding to the strategy played, by computing the average between the previous perception and the payoff received using the number of times that each strategy has been played. The resultant process turns out to be a variation of the one studied by Cominetti et al. [9], but in our case, players use more information on the history of play. Using the tools provided by the stochastic approximation theory (see e.g., Benaïm [3], Benveniste et al. [4], Kushner and Yin [17]), the asymptotic behavior of the process can be analyzed by studying a related continuous dynamics. We are interested in the case in which players use the Logit decison rule and our aim is to find general conditions to have almost sure, or with positive probability, convergence to an attractor of the associated ODE. This case is particularly interesting because the rest points of the ODE are the Nash equilibria of a related game.

This paper is organized as follows. Section 2 describes the very basic aspects of the stochastic approximation theory. Section 3 states precisely our model in the framework of an infinitely repeated N-player normal form game. In Section 4, we restate our algorithm to make it fit in the set-up of the stochastic approximation and we provide a general almost sure convergence result. In Section 5.1 we treat the case of the Logit rule in detail. We start by finding an explicit condition to ensure almost sure convergence derived from Section 4. This condition demands the smoothing parameters associated to the Logit rule to be sufficiently small. It is worth noting that, up to this point, we proved that the same results obtained for the process studied by Cominetti et al. [9] hold in our setting. Given this fact, we perform a comparison between these two processes in terms of the path-wise rate of convergence. Later, under a weaker assumption, we study the convergence with

positive probability to attractors. We apply this result to a particular traffic game on a simple network (studied as an application in [9]) showing that a convergence with positive probability property holds under a much weaker assumption than in the general case. Next, we provide some examples where the convergence is lost. Finally, in Section 6, we prove that part of the analysis can be recovered if players face random payoffs. For that purpose, we use some known techniques that, to our knowledge, do not seem very exploited in the framework of learning in games.

2 Preliminaries

In this section, we revisit some basic aspects of the stochastic approximation theory following the approach in Benaïm [3]. The motivation is to study the following discrete process in $\mathbb{R}^d$

$$z_{n+1} - z_n = \gamma_{n+1}\big(H(z_n) + V_{n+1}\big), \tag{2.1}$$

where $(\gamma_n)_n$ is a nonnegative step-size sequence, $H : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous function and $(V_n)_n$ is a (deterministic or random) noise. Let us denote by $\mathcal{L}(z_n)$ the limit set of the sequence $(z_n)_n$, i.e., the set of points z such that $\lim_{l\to+\infty} z_{n_l} = z$ for some sequence $n_l \to +\infty$.

The connection between the asymptotic behavior of the discrete process (2.1) and the asymptotic behavior of the continuous dynamics

$$\dot{z} = H(z) \tag{2.2}$$

is obtained as follows. Given $\varepsilon > 0$, $T > 0$, a set $Z \subseteq \mathbb{R}^d$ and two points $x, y \in Z$, we say that there is an (ε, T)-chain in Z between x and y if there exist k solutions of (2.2) $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ and times $\{t_1, \dots, t_k\}$ greater than T such that

(1) $\mathbf{x}_i([0, t_i]) \subseteq Z$ for all $i \in \{1, \dots, k\}$,
(2) $\|\mathbf{x}_i(t_i) - \mathbf{x}_{i+1}(0)\| < \varepsilon$ for all $i \in \{1, \dots, k-1\}$,
(3) $\|\mathbf{x}_1(0) - x\| < \varepsilon$ and $\|\mathbf{x}_k(t_k) - y\| < \varepsilon$.

Definition 2.1. *A set $D \subseteq \mathbb{R}^d$ is* Internally Chain Transitive (ICT) *for the dynamics* (2.2) *if it is compact and for all $\varepsilon > 0$, $T > 0$ and $x, y \in D$ there exists an (ε, T)-chain in D between x and y.*

This definition is derived from the notion of *Internally Chain Recurrent* sets introduced by Conley [10]. Roughly speaking, on an ICT set, we can link any two points by a chain of solutions of the dynamics (2.2) by allowing small perturbations. ICT sets are compact, invariant and attractor-free. In Benaïm [3] the following general theorem is proved.

Theorem 2.2. *Consider the discrete process* (2.1). *Assume that H is a Lipschitz function and that*

- *(a) the sequence (γ_n) is deterministic, $\gamma_n \geq 0$, $\sum_n \gamma_n = +\infty$ and $\gamma_n \to 0$,*
- *(b) $\sup_{n\in\mathbb{N}} \|z_n\| < +\infty$, and*
- *(c) for any $T > 0$*

$$\lim_{n\to+\infty} \sup\left\{ \left\| \sum_{i=n}^{k-1} \gamma_{i+1} V_{i+1} \right\| ; k \in \{n+1, \dots, m(\sum_{j=1}^{n} \gamma_j + T)\} \right\} = 0,$$

where $m(t)$ is the largest integer l such that $t \geq \sum_{j=1}^{l} \gamma_j$. Then $\mathcal{L}(z_n)$ is an ICT *set for the dynamics* (2.2).

Remark 2.3. In the case where the noise $(V_n)_n$ in (2.1) is a martingale difference sequence with respect to some filtration on a probability space, we say that (2.1) is a Robbins–Monro [23] algorithm. In this framework if, for instance, $\sup_n \mathbb{E}(\|V_n\|^2) < +\infty$ and $(\gamma_n)_n \in l^2(\mathbb{N})$ then assumption (c) in Theorem 2.2 holds with probability one (see Benaïm [3, Proposition 4.2]). Moreover this result is still valid if the noise can be decomposed into a martingale difference process plus a random variable that converges to zero almost surely.

3 The model

An N-player normal form game is introduced as follows. Let $A = \{1, 2, \ldots, N\}$ be the set of players. For every $i \in A$ let S^i be the finite strategy set for player i and let the set $\Delta^i = \{z \in \mathbb{R}^{|S^i|}; z^i \geq 0, \sum_i z^i = 1\}$ denote her mixed strategy set. $S = \prod_{i \in A} S^i$ is the set of strategy profiles and $\Delta = \prod_{i \in A} \Delta^i$ is the set of mixed strategy profiles. We write as $(s, s^{-i}) \in S$ the strategy profile where player i uses the strategy $s \in S^i$ and her opponents use the strategy profile $s^{-i} \in \prod_{j \neq i} S^j$ and we adopt the same notation when a mixed strategy profile is involved. The payoff function to each player $i \in A$ is denoted by $G^i : S \to \mathbb{R}$ as well as its multilinear extension $G^i : \Delta \to \mathbb{R}$.

The game is repeated infinitely and we suppose that players are not informed about the structure of the game, i.e., neither the number of players (or their strategies) nor the payoff functions are known. At the stage $n \in \mathbb{N}$, each player i selects a strategy $s^i_n \in S^i$ using the mixed strategy $\sigma^i_n \in \Delta^i$. Then, she gets her own payoff $g^i_n = G^i(s^i_n, s^{-i}_n)$ and this is the only information she receives.

For every $n \in \mathbb{N}$ and for each player i, we assume that the mixed strategy at stage n , $\sigma^i_n \in \Delta^i$, is determined as a function of a previous *perception vector* $x^i_{n-1} \in \mathbb{R}^{|S^i|}$, i.e., $\sigma^i_n = \sigma^i(x^i_{n-1})$ with $\sigma^i : \mathbb{R}^{|S^i|} \to \Delta^i$. The state space for the perception vector profiles $x = (x^1, \ldots, x^N) \in \prod_{i \in A} \mathbb{R}^{|S^i|}$ is denoted X. We also assume that, for every $i \in A$,

$$\begin{aligned} &\text{the function } \sigma^i : \mathbb{R}^{|S^i|} \to \Delta^i \text{ is continuous, and} \\ &\text{for all } s \in S^i \text{ and } x^i \in \mathbb{R}^{|S^i|},\ \sigma^{is}(x^i) > 0. \end{aligned} \tag{A}$$

We refer to the function $\sigma : X \to \Delta$ with $\sigma(x) = (\sigma^1(x^1), \ldots, \sigma^N(x^N))$ as the *decision rule* of the players.

At the end of the stage n, each player i uses the value g^i_n and x^i_{n-1} to obtain the new perception vector x^i_n, and so on. The manner in which x_n is updated is called the *updating rule* of the players.

Cominetti et al. [9] study the following updating rule

$$x^{is}_{n+1} = \begin{cases} \big(1 - \gamma_{n+1}\big) x^{is}_n + \gamma_{n+1} g^i_{n+1}, & \text{if } s = s^i_{n+1}, \\ x^{is}_n, & \text{otherwise,} \end{cases} \tag{3.1}$$

where we assume that $\gamma_n = \frac{1}{n}$ (see Remark 4.3 for an explanation on this choice).

In this paper we consider a variation of (3.1). Players will use more information by taking into account the number of times that their actions have been played. Explicitly, we define the adjusted

process (AP) by

$$x_{n+1}^{is} = \begin{cases} \left(1 - \dfrac{1}{\theta_n^{is}}\right) x_n^{is} + \dfrac{1}{\theta_n^{is}} g_{n+1}^i, & \text{if } s = s_{n+1}^i, \\ x_n^{is}, & \text{otherwise,} \end{cases} \tag{AP}$$

where θ_n^{is} denotes the number of times that strategy s has been used by player $i \in A$ up to time n. Given the particular structure in (AP), we can suppose that x_n lies in a compact subset of X for all $n \in \mathbb{N}$. We also note that (A) implies that the decision rule can be assumed component-wise bounded away from zero.

For feasibility issues we suppose that $\theta_0^{is} > 0$ for all $i \in A$ and $s \in S^i$ and then θ_n^{is} is not exactly the number of times that the action $s \in S^i$ has been used by player i (since it contains the initial condition), but we keep this interpretation. This assumption is not relevant in the asymptotic analysis developed further on.

As usual we denote by $\mathcal{F}_n$ the σ-algebra generated by the history up to time n, $\mathcal{F}_n = \sigma\big((s_m, g_m)_{1 \le m \le n}\big)$, where $s_m = (s_m^1, \ldots, s_m^N)$ and $g_m = (g_m^1, \ldots, g_m^N)$.

4 Asymptotic analysis

The main difficulty to analyze (AP) using the tools decribed in Section 2, is that we have a stochastic algorithm in discrete time where the step-size is random and, moreover, depends on the coordinates of the vector to update. Thus, in order to study the asymptotic properties of our adaptive process, let us restate the updating scheme (AP) in the following manner

$$\begin{aligned} x_{n+1}^{is} - x_n^{is} &= \frac{1}{\theta_n^{is}} (g_{n+1}^i - x_n^{is}) \mathbb{1}_{\{s = s_{n+1}^i\}} \\ &= \frac{1}{(n+1)\lambda_n^{is}} (g_{n+1}^i - x_n^{is}) \mathbb{1}_{\{s = s_{n+1}^i\}}, \end{aligned}$$

where $\lambda_n^{is} = \frac{\theta_n^{is}}{n+1}$ is interpreted as the empirical frequency of action s for player i up to time n and $\mathbb{1}_C$ stands for the indicator function of the set C. Without loss of generality, assume that $0 < \theta_0^{is} \le 1$, for every $i \in A$ and $s \in S^i$, in order to have $\lambda^i \in \Delta^i$ for al $n \in \mathbb{N}$. Standard computations involving averages show that

$$\lambda_{n+1}^{is} - \lambda_n^{is} = \frac{1}{n+1} \big(\mathbb{1}_{\{s = s_{n+1}^i\}} - \lambda_n^{is} + b_{n+1} \big),$$

and

$$b_{n+1} = -\frac{1}{n+2} \big(\mathbb{1}_{\{s = s_{n+1}^i\}} - \lambda_n^{is} \big) = O\left(\frac{1}{n}\right). \tag{4.1}$$

Then we can express (AP) differently by introducing the empirical frequency of play. The new form is the (up to a vanishing term) martingale difference scheme

$$\begin{cases} x_{n+1}^{is} - x_n^{is} = \dfrac{1}{n+1} \Big[\dfrac{\sigma^{is}(x_n^i)}{\lambda_n^{is}} (G^i(s, \sigma^{-i}(x_n)) - x_n^{is}) + U_{n+1}^{is} \Big], \\ \\ \lambda_{n+1}^{is} - \lambda_n^{is} = \dfrac{1}{n+1} \big[\sigma^{is}(x_n^i) - \lambda_n^{is} + M_{n+1}^{is} \big], \end{cases} \tag{APD}$$

where the noise terms are explicitly

$$
\begin{aligned}
U_{n+1}^{is} &= \frac{1}{\lambda_n^{is}}(g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} - \left[\frac{\sigma^{is}(x_n^i)}{\lambda_n^{is}}(G^i(s,\sigma^{-i}(x_n)) - x_n^{is})\right],\\
&= \frac{1}{\lambda_n^{is}}(g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} - \mathbb{E}\big(\frac{1}{\lambda_n^{is}}(g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} \,|\, \mathcal{F}_n\big),\\
M_{n+1}^{is} &= \mathbb{1}_{\{s=s_{n+1}^i\}} - \sigma^{is}(x_n^i) + b_{n+1},\\
&= \big(\mathbb{1}_{\{s=s_{n+1}^i\}} - \lambda_n^{is}\big) - \mathbb{E}\big(\big(\mathbb{1}_{\{s=s_{n+1}^i\}} - \lambda_n^{is}\big) \mid \mathcal{F}_n\big) + b_{n+1}.
\end{aligned}
\tag{4.2}
$$

From now on, we denote by $\epsilon_n = (U_n, M_n)$ the noise term associated to our process.

The scheme (APD) will allow us to deal with the random (and player-dependent) character of the step-size in (AP). Now, in the spirit of Theorem 2.2, the asymptotic behavior of (APD) is related with the continuous dynamics

$$
\begin{cases}
\dot{x}_t^{is} = \dfrac{\sigma^{is}(x_t^i)}{\lambda_t^{is}}\Big(G^i(s,\sigma^{-i}(x_t)) - x_t^{is}\Big) & = \Psi_x^{is}(x_t,\lambda_t),\\[2ex]
\dot{\lambda}_t^{is} = \sigma^{is}(x_t^i) - \lambda_t^{is} & = \Psi_\lambda^{is}(x_t,\lambda_t),
\end{cases}
\tag{4.3}
$$

with $\Psi_x : X \times \Delta \to \prod_{i\in A}\mathbb{R}^{|S^i|}$ and $\Psi_\lambda : X \times \Delta \to \prod_{i\in A}\Delta_0^i$, and Δ_0^i standing for the tangent space to Δ^i, i.e., $\Delta_0^i = \{z \in \mathbb{R}^{|S^i|};\, \sum_{s\in S^i} z^s = 0\}$. Let us denote Ψ the function defined by $\Psi(x,\lambda) = (\Psi_x(x,\lambda), \Psi_\lambda(x,\lambda))$.

For the sake of completeness, let us write the process (3.1) as

$$
x_{n+1}^{is} - x_n^{is} = \frac{1}{n+1}\big[\sigma^{is}(x_n^i)(G^i(s,\sigma^{-i}(x_n)) - x_n^{is}) + \tilde{U}_{n+1}^{is}\big],
\tag{4.4}
$$

with noise term given by

$$
\begin{aligned}
\tilde{U}_{n+1}^{is} &= (g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} - \sigma^{is}(x_n^i)(G^i(s,\sigma_n^{-i}(x_n)) - x_n^{is}),\\
&= (g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} - \mathbb{E}((g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} \mid \mathcal{F}_n).
\end{aligned}
\tag{4.5}
$$

Therefore, the corresponding continuous dynamics is given by

$$
\dot{x}_t^{is} = \sigma^{is}(x_t^i)\big(G^i(s,\sigma^{-i}(x_t)) - x_t^{is}\big) = \Phi^{is}(x_t),
\tag{4.6}
$$

where $\Phi : X \to \prod_{i\in A}\mathbb{R}^{|S^i|}$.

Remark 4.1. Observe that the following simple fact holds

$$
(x,\sigma(x)) \in X \times \Delta \text{ is a rest point of (4.3)} \Leftrightarrow x \in X \text{ is a rest point of (4.6)}.
$$

In the sequel we will show that asymptotic properties similar to those of (4.4) can be obtained for our process. This means that explicit conditions can be found to ensure that the process (APD) converges almost surely to a global attractor for the dynamics (4.3).

Recall that we have assumed that, for every $n \in \mathbb{N}$ and $i \in A$, the mixed action $\sigma_n^i \in \Delta^i$ is component-wise bounded away from zero. The purpose of the next simple lemma is to verify that the same holds, almost surely, for the empirical frequencies of play.

Lemma 4.2. *For $n \geq 1$, let σ_n be a probability distribution over a finite set T and let i_{n+1} be an element of T which is drawn with law σ_n and assume $(\sigma_n)_n$ is adapted to the natural filtration generated by the history. For all $j \in T$, set*

$$\lambda_n^j = \gamma_n \sum_{p=1}^{n} \mathbb{1}_{\{i_p = j\}},$$

where $(\gamma_n)_n$ is a decreasing positive real sequence such that $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < +\infty$. Assume that there exists $\overline{\sigma} > 0$ such that $\sigma_n^j \geq \overline{\sigma}$. Then

$$\liminf_{n \to +\infty} \lambda_n^j \geq \overline{\sigma},$$

almost surely, for every $j \in T$.

Proof. Fix $j \in T$ and let $\mathcal{F}_k$ be the σ-algebra generated by the history $\{i_1, \dots, i_k\}$ up to time k. Then we have that $\mathbb{E}(\mathbb{1}_{\{i_k=j\}} \mid \mathcal{F}_{k-1}) = \sigma_{k-1}^j \geq \overline{\sigma}$. On the other hand the random process $(\phi_n^j)_n$ given by

$$\phi_n^j = \sum_{k=1}^{n} \gamma_k \big(\mathbb{1}_{\{i_k=j\}} - \mathbb{E}(\mathbb{1}_{\{i_k=j\}} \mid \mathcal{F}_{k-1}) \big)$$

is a martingale and $\sup_{n \in \mathbb{N}} (\phi_n^j)^2 \leq C \cdot \sum_{p \geq 1} \gamma_p^2 < +\infty$ for some constant C. Hence $(\phi_n^j)_n$ converges almost surely. Now Kronecker's lemma (see e.g., Shiryaev [25, Lemma IV.3.2]) gives that

$$\lim_{n \to +\infty} \gamma_n \sum_{k=1}^{n} \big(\mathbb{1}_{\{i_k=j\}} - \mathbb{E}(\mathbb{1}_{\{i_k=j\}} \mid \mathcal{F}_{k-1}) \big) = 0. \tag{4.7}$$

So that $\gamma_n \sum_{k=1}^{n} (\mathbb{1}_{\{i_k=j\}} - \mathbb{E}(\mathbb{1}_{\{i_k=j\}} \mid \mathcal{F}_{k-1})) \leq \lambda_n^j - \overline{\sigma}$. Taking $\liminf$ and using (4.7) we conclude. □

Remark 4.3. Observe that we have defined an adjusted version of (4.4) with $\gamma_n = \frac{1}{n}$. The previous lemma shows that all the analysis in this section can be carried out for more general step-size sequences $(\gamma_n)_n$ by taking $\lambda_n^{is} = \theta_n^{is} \gamma_n$. We adopt $\gamma_n = \frac{1}{n}$ to keep the convenient interpretation in terms of frequencies of play.

Proposition 4.4. *The process* (APD) *converges almost surely to an* ICT *set for the continuous dynamics* (4.3).

Proof. We only have to show that our process satisfies the hypotheses of Theorem 2.2. The assumptions concerning the regularity of the function involved, the step-size sequence and the boundedness of the process $(x_n, \lambda_n)_n$ hold immediately.

According to (4.2), $(M_n)_n$ is almost surely bounded and can be written as a martingale difference scheme plus a vanishing term. Observe that $\mathbb{E}(U_{n+1} \mid \mathcal{F}_n) = 0$ and that

$$\left\| U_{n+1}^{is} \right\| \leq C / \lambda_n^{is},$$

for some constant C. Then Lemma 4.2 implies that U_n is almost surely bounded. In view of Remark 2.3, assumption (c) of Theorem 2.2 holds for the noise term $\epsilon_n = (U_n, M_n)$ and the conclusion follows. □

Let us define the function $F : X \to \prod_i \mathbb{R}^{|S^i|}$ by

$$F^{is}(x) = G^i(s, \sigma^{-i}(x)). \tag{4.8}$$

Cominetti et al. [9] show that if the function F is contracting for the infinity norm, then the process (4.4) converges almost surely to the unique rest point of the dynamics (4.6). The following result shows that the same holds for the process (APD) by adding a slighlty stronger assumption on the decision rule σ.

Proposition 4.5. *Assume that F is contracting for the infinity norm and that, for every $i \in A$, the function σ^i is Lipschitz for the infinity norm. Then there exists a unique rest point $(x_*, \sigma(x_*)) \in X \times \Delta$ of* (4.3)*. Furthermore, the set $\{(x_*, \sigma(x_*))\}$ is a global attractor and the process* (APD) *converges almost surely to $(x_*, \sigma(x_*))$.*

Proof. According to Remark 4.1, $(x_*, \sigma(x_*)) \in X \times \Delta$ is a rest point of (4.3) if and only if $F(x_*) = x_*$, hence the existence and uniqueness follow from the fact that F is contracting.

Let $0 \leq L < 1$ and K_i be the Lipstchitz constants associated to the functions F and σ^i, $i \in A$, respectively. We want to find a suitable strict Lyapunov function, i.e., a function V that decreases along the solution paths and that verifies $V^{-1}(\{0\}) = \{(x_*, \lambda_*)\}$ with $\lambda_* = \sigma(x_*)$. Let $V : X \times \Delta \to \mathbb{R}_+$ be defined by

$$V(x, \lambda) = \max\left\{ \|x - x_*\|_\infty , \frac{1}{\zeta} \|\lambda - \lambda_*\|_\infty \right\},$$

where $\zeta > 0$ will be defined later on. The function V is the maximum of a finite number of smooth functions, therefore it is absolutely continuous and its derivatives are the evaluation of the derivatives of the function attaining the maximum. We distinguish two cases:

Case 1. $V(x_t, \lambda_t) = \|x_t - x_*\|_\infty$. Let $i \in A$ and $s \in S^i$ be such that $V(x_t, \lambda_t) = |x_t^{is} - x_*^{is}|$. Let us assume that $x_t^{is} - x_*^{is} \geq 0$. Then, for almost all $t \in \mathbb{R}$,

$$\begin{aligned}
\frac{d}{dt} V(x_t, \lambda_t) &= \frac{d}{dt}(x_t^{is} - x_*^{is}) \\
&= \frac{\sigma^{is}(x_t^i)}{\lambda_t^{is}} \big(F^{is}(x_t) - F^{is}(x_*) + x_*^{is} - x_t^{is} \big) \\
&\leq -\xi(1 - L) \|x_t - x_*\|_\infty \\
&= -\xi(1 - L) V(x_t, \lambda_t),
\end{aligned}$$

for some $\xi > 0$ such that $\sigma^{is}(x) \geq \xi$ for every $i \in A$ and $s \in S^i$. If $x_t^{is} - x_*^{is} < 0$, the computations are analogous.

Case 2.: $V(x_t, \lambda_t) = \frac{1}{\zeta} \|\lambda_t - x_*\|_\infty$. Let $i \in A$ and $s \in S^i$ be such that $V(x_t, \lambda_t) = \frac{1}{\zeta}|\lambda_t^{jr} - \lambda_*^{jr}|$. We also assume that $\lambda_t^{jr} - \lambda_*^{jr} \geq 0$. Then, for almost all $t \in \mathbb{R}$,

$$\begin{aligned}
\frac{d}{dt} V(x_t, \lambda_t) &= \frac{1}{\zeta} \big[\sigma^{jr}(x_t^j) - \sigma^{jr}(x_*) + \lambda_*^{jr} - \lambda_t^{jr} \big] \\
&\leq -\frac{1}{\zeta} \|\lambda_t - \lambda_*\|_\infty + \frac{1}{\zeta} |\sigma^{jr}(x_t^j) - \sigma^{jr}(x_*)| \\
&\leq -V(x_t, \lambda_t) + \frac{\max_i K_i}{\zeta} \|x_t - x_*\|_\infty \\
&= -(1 - \frac{\max_i K_i}{\zeta}) V(x_t, \lambda_t),
\end{aligned}$$

and we take $\zeta > 0$ sufficiently large to have $1 > \max_i K_i/\zeta$. Again, if the relation $\lambda_t^{jr} - \lambda_*^{jr} < 0$ holds, the computations are the same.

Hence $V(x_t, \lambda_t) \leq -KV(x_t, \lambda_t)$ for some $K > 0$. So V decreases exponentially fast along the solution paths of the dynamics and $V(x, \lambda) = 0$ if and only if $(x, \lambda) = (x_*, \lambda_*)$. Therefore the set $\{(x_*, \lambda_*)\}$ is a global attractor which is the unique ICT set for (4.3) (see [3, Corollary 5.4]). Proposition 4.4 finishes the proof. □

5 Logit rule

In this section we will focus on the analysis of a particular decision rule: the Logit rule which has a large foundation in the field of discrete choice models as well as in game theory. Explicitly the decision rule $\sigma : X \to \Delta$ is given by

$$\sigma^{is}(x^i) = \frac{\exp\left(\beta_i x^{is}\right)}{\sum\limits_{r \in S^i} \exp\left(\beta_i x^{ir}\right)}, \tag{5.1}$$

for every $i \in A$ and $s \in S^i$, where $\beta_i > 0$ is called the smoothing parameter for player i. According to Remark 4.1, the following result shows that the rest points of the dynamics (4.3) are the Nash equilibria for an entropy perturbed version of the original game.

Lemma 5.1 (Cominetti et al. [9]). *Under the Logit decision rule* (5.1)*, if $x \in X$ is a rest point of the dynamics* (4.6)*, then $\sigma(x)$ is a Nash equilibrium of a game where the strategy set for the each player i is Δ^i and her payoff $\overline{G}^i : \Delta \to \mathbb{R}$ is given by*

$$\overline{G}^i(\pi) = \sum_{s \in S^i} \pi^{is} G^i(s, \pi^{-i}) - \frac{1}{\beta_i} \sum_{s \in S^i} \pi^{is}\big(\ln(\pi^{is}) - 1\big). \tag{5.2}$$

5.1 Almost sure convergence

We want to apply Proposition 4.5 in this particular framework. For that purpose, let us introduce the maximal unilateral deviation payoff that a single player can face,

$$\eta = \max_{\substack{i \in A, s \in S^i \\ r_1, r_2 \in \tilde{S}^{-i}}} |G^i(s, r_1) - G^i(s, r_2)|, \tag{5.3}$$

where $\tilde{S}^{-i} = \{(r_1, r_2) \in S^{-i} \times S^{-i}; r_1^k \neq r_2^k \text{ for exactly one } k\}$. Now we can show the following proposition that ensures that, if the parameters are sufficiently small, the unique attractor is attained with probability one. From now on, we denote $\alpha = \max_{i \in A} \sum_{j \neq i} \beta_j$.

Proposition 5.2. *If $2\eta\alpha < 1$, the discrete process* (APD) *converges almost surely to the unique rest point $(x_*, \sigma(x_*))$ of the dynamics* (4.3)*.*

Proof. We know from Cominetti et al. [9, Proposition 5] that, if $2\eta\alpha < 1$, the function F (defined in (4.8)) is contracting for the infinity norm. Observe also that, for every $i \in A$, the function σ^i is Lipschitz for the infinity norm, since it is a smooth function defined on a compact set. Therefore, Proposition 4.5 applies. □

Rate of Convergence

Up to this point, we were able to reproduce some of the theoretical results of the original model (4.4) regarding its almost sure convergence to global attractors. Now, we want to justify the inclusion of a counter of the previous actions in terms of the rate of convergence when both learning processes (APD) and (4.4) converge almost surely to (x_*, λ_*) and x_*, respectively, and the step-size $\gamma_n = \frac{1}{n}$ is considered. This rate of convergence is deeply linked with the largest real part eigenvalue of the Jacobian matrix of the functions $\Psi = (\Psi_x, \Psi_\lambda)$ and Φ at the respective rest points.

Let us denote $\rho(\mathcal{B})$ the maximal real part of the eigenvalues of a matrix $\mathcal{B} \in \mathbb{R}^{k\times k}$, i.e.,

$$\rho(\mathcal{B}) = \max\{\mathrm{Re}(\mu_j);\ j = 1, \dots, k,\ \text{where } \mu_j \in \mathbb{C} \text{ is an eigenvalue of the matrix } \mathcal{B}\}.$$

We say that a matrix $\mathcal{B}$ is stable if $\rho(\mathcal{B}) < 0$.

Lemma 5.3. *Assume that $2\eta\alpha < 1$. Let (x_*, λ_*) and x_* be the unique rest points of the dynamics (4.3) and (4.6), respectively. Then*

$$-1 \le \rho(\nabla\Psi(x_*, \lambda_*)) < -\frac{1}{2} \le -\frac{N}{\sum\limits_{k\in A} |S^k|} \le \rho(\nabla\Phi(x_*)) < 0. \tag{5.4}$$

Proof. Straigthforward computations concerning the function Ψ (see (4.3)) show that

$$\frac{\partial\Psi_\lambda^{is}}{\partial x^{jr}}(x_*, \lambda_*) = 0 \quad \text{and} \quad \frac{\partial\Psi_\lambda^{is}}{\partial \lambda^{jr}}(x_*, \lambda_*) = -\mathbb{1}_{\{is=jr\}},$$

for every $i, j \in A$ and $(s, r) \in S^i \times S^j$. Therefore, the matrix $\nabla\Psi(x_*, \lambda_*)$ looks like

$$\nabla\Psi(x_*, \lambda_*) = \begin{pmatrix} \nabla_x\Psi_x(x_*, \lambda_*) & 0 \\ L & -I \end{pmatrix}, \tag{5.5}$$

where I stands for the identity matrix and $\nabla_x\Psi_x(x_*, \lambda_*)$ denotes the Jacobian matrix of Ψ_x with respect to x at (x_*, λ_*). Notice that the interesting eigenvalues of this matrix are given by its upper-left block because of the zero block and the identity matrix on the right side in (5.5). Observe also that $\frac{\partial\Psi_x^{is}}{\partial x^{is}}(x_*, \lambda_*) = -1$, i.e., the matrix $\nabla_x\Psi_x(x_*, \lambda_*)$ has diagonal terms equal to -1.

On the other hand, we know that every eigenvalue of a complex matrix $\mathcal{B} = (\mathcal{B}_{pq})$ lies within at least one of the Gershgorin discs $D_p(\mathcal{B}) = \{z \in \mathbb{C}, |z - \mathcal{B}_{pp}| \le R_p\}$ where $R_p = \sum_{q\neq p} |\mathcal{B}_{pq}|$. Given the specific form of the matrix $\nabla_x\Psi(x_*, \lambda_*)$ we can estimate the position of its eigenvalues. So, in our case,

$$R_{is} = \sum_{\substack{j\in A,\\ j\neq i}} \sum_{r\in S^j} \left| \frac{\partial\Psi_x^{is}}{\partial x^{jr}}(x_*, \lambda_*) \right|,$$

since $\frac{\partial\Psi_x^{is}}{\partial x^{jr}}(x_*, \lambda_*) = 0$ if $i = j$ and $r \neq s$. This follows from the fact that $F^{is}(x)$ (defined in (4.8)) is independent of the vector x^i. Explicitly,

$$\frac{\partial\Psi_x^{is}}{\partial x^{jr}}(x_*, \lambda_*) = \beta_j \sigma_*^{jr} \left[G^i(s, r, \sigma_*^{-(i,j)}) - G^i(s, \sigma_*^{-i}) \right],$$

where

$$G^i(s,r,\sigma_*^{-(i,j)}) = \sum_{\substack{a\in S^{-i}\\ a^j=r}} G^i(s,a)\prod_{\substack{k\neq i\\ k\neq j}} \sigma_*^{ka^k},$$

for $i \neq j$. So that

$$\begin{aligned} R_{is} &= \sum_{\substack{j\in A\\ j\neq i}} \beta_j \sum_{r\in S^j} \sigma_*^{jr} \left| G^i(s,r,\sigma_*^{-(i,j)}) - G^i(s,\sigma_*^{-i}) \right| \\ &\leq \eta\alpha. \end{aligned}$$

Then we have that all the eigenvalues of the matrix $\nabla_x \Psi_x(x_*,\lambda_*)$ are contained in the complex disc

$$\{z\in\mathbb{C}, |z+1| \leq \eta\alpha\} \supseteq \bigcup_{\substack{i\in A\\ s\in S^i}} D_{is}(\nabla_x\Psi_x(x_*,\lambda_*)), \tag{5.6}$$

which implies that $\rho(\nabla\Psi(x_*,\lambda_*)) < -1/2$.

Analogous computations involving the function Φ show that

$$D_{is}(\nabla\Phi(x_*)) \subseteq \{z\in\mathbb{C}, |z+\sigma_*^{is}| \leq \sigma_*^{is}\eta\alpha\},$$

for every $i \in A$ and $s \in S^i$. Since $-\sigma_*^{is} + \sigma_*^{is}\eta\alpha < 0$, then $\rho(\nabla\Phi(x_*)) < 0$.

The fact that $-1 \leq \rho(\nabla\Psi(x_*,\lambda_*))$ is evident. Finally, the inequality $-N/\sum_k |S^k| \leq \rho(\nabla\Phi(x_*))$ follows since the trace of the matrix $\nabla\Phi(x_*)$ is equal to $-N$. ☐

Remark 5.4. Notice that $1/2 = N/\sum_k |S^k|$ if and only if $|S_k| = 2$ for all $k\in A$.

The following reduced version of Chen [8, Theorem 3.1.1] will be useful.

Theorem 5.5. *Consider the discrete process given by* (2.1)*. Assume that the following hold.*

(a) *For every $n\in\mathbb{N}$, $\gamma_n > 0$, $\lim_{n\to+\infty}\gamma_n = 0$, $\sum_n \gamma_n = +\infty$ and*

$$\lim_{n\to+\infty} \frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}\gamma_n} = \overline{\gamma} \geq 0.$$

(b) *$z_n \to z_0$ almost surely.*

(c) *There exists $\delta\in(0,1]$ such that*

 (c.1) *for a path such that $z_n\to z_0$, the noise V_n can be decomposed in $V_n = V_n' + V_n''$ where*

$$\sum_{n\geq 1} \gamma_n^{1-\delta} V_{n+1}' < +\infty \quad \text{and} \quad V_n'' = O(\gamma_n^\delta),$$

 (c.2) *the function H is locally bounded and is differentiable at z_0 such that $H(z) = \overline{H}(z-z_0) + r(z)$ where $r(z_0) = 0$ and $r(z) = o(\|z-z_0\|)$ as $z\to z_0$ and*

 (c.2) *the matrix $\overline{H}$ is stable and, furthermore, $\overline{H} + \delta\overline{\gamma} I$ is also stable.*

Then, almost surely,

$$(1/\gamma_n)^\delta (z_n - z_0) \to 0, \quad \text{as } n \to +\infty.$$

The previous result allows us to show that, in some sense, our algorithm is faster. This will show that, under the common hypothesis $2\eta\alpha < 1$ (which ensures almost sure convergence for both processes), the employment of the adjusted process (APD) will help the players to adapt faster their behavior than the original process (4.4).

Proposition 5.6. *Assume that $2\eta\alpha < 1$ and let $(x_*, \lambda_*) \in X \times \Delta$ and $x_* \in X$ be the unique rest points of the dynamics (4.3) and (4.6), respectively. Then the following estimates hold*

(i) *for almost all trajectories of* (4.4)

$$n^\delta (x_n - x_*) \to 0, \quad \text{as } n \to +\infty,$$

for every $\delta \in (0, |\rho(\nabla\Phi(x_*))|)$,

(ii) *for almost all trajectories of* (APD)

$$n^\delta \big((x_n, \lambda_n) - (x_*, \lambda_*)\big) \to 0, \quad \text{as } n \to +\infty,$$

for every $\delta \in (0, 1/2)$.

Proof. Recall that $\epsilon_n = (U_n, M_n)$ and $\tilde{U}_n$ are the noise terms associated to (APD) and (4.4), respectively (see (4.2) and (4.5)). We observe that, for both processes, hypotheses (a) and (b) in Theorem 5.5 are immediately satisfied since $\gamma_n = \frac{1}{n}$, (with $\overline{\gamma} = 1$) and since Proposition 5.2 applies. Let us verify that condition (c) holds.

(i) Fix $\delta \in (0, |\rho(\nabla\Phi(x_*))|)$. The random process $(\tilde{U}_n)_n$ is almost surely bounded and satisfies that $\mathbb{E}(\tilde{U}_{n+1} \mid \mathcal{F}_n) = 0$. Therefore, $Z_n = \sum_{k=1}^{n} (1/k)^{1-\delta} \tilde{U}_{k+1}$ is a martingale where $\sup_n \|Z_n\|^2 < \sum_{k=1}^{+\infty} (1/k)^{2(1-\delta)} < +\infty$, thus convergent (since $\delta < 1/2$). To conclude, observe that the function Φ is smooth and that the matrix $\nabla\Phi(x_*) + \delta I$ is stable.

(ii) Fix $\delta \in (0, 1/2)$. We repeat the argument by noticing that $\epsilon_n = \tilde{\epsilon}_n + \tilde{b}_n$ where $\tilde{b}_n = O(1/n)$ and $\mathbb{E}(\tilde{\epsilon}_{n+1} \mid \mathcal{F}_n) = 0$. To finish, we use the fact that the matrix $\nabla\Psi(x_*, \lambda_*) + \delta I$ is stable since inequality (5.4) holds.

□

Remark 5.7. Two important comments are in order.

(a) For the process (APD), if the matrix $C_n = \mathbb{E}(\epsilon_{n+1}^T \epsilon_{n+1} \mid \mathcal{F}_n)$ converges almost surely to a deterministic positive definite matrix C, then $\sqrt{n}((x_n, \lambda_n) - (x_*, \lambda_*))$ converges in distribution to a normal random variable (see e.g., Benveniste et al. [4] or Kushner and Yin [17]). For the process (4.4), if the sequence $\tilde{C}_n = \mathbb{E}(\tilde{U}_{n+1}^T \tilde{U}_{n+1} \mid \mathcal{F}_n)$ converges almost surely to a deterministic positive definite matrix $\tilde{C}$, then it can be shown (see Duflo [11]) that $n^{|\rho(\nabla\Phi(x_*))|}(x_n - x_*)$ converges almost surely to a finite random variable. For instance, if we consider the game defined by (5.7), we can show that both sequences $(C_n)_n$ and $(\tilde{C}_n)_n$ converge almost surely to a deterministic positive definite matrix and that $|\rho(\nabla\Phi(x_*))| < 1/2$. Therefore, in general, nothing more can be said when the step-size $\gamma_n = \frac{1}{n}$ is considered. Figure 1 consists in the results of a numerical experience in this particular example where $2\eta\alpha = 0.8$ and $|\rho(\nabla\Phi(x_*))| \approx 0.3$.

(b) Observe that a better rate can be achieved for the process (4.4) if the step-size is given by $\gamma_n = \frac{a}{n}$ for $a > |\rho(\nabla\Psi(x_*))|$. This leads to the rate $o(n^{-\delta})$ for all $\delta \in (0, 1/2)$. However it is fairly unrealistic to assume that the players know in advance this information. Nevertheless, we always have that $|\rho(\nabla\Phi(x_*))| < |\rho(\nabla\Psi(x_*, \lambda_*))|$ and then the scheme (APD) can reach at least the same path-wise rate of convergence under the hypotheses of Proposition 5.6 and independent of the step-size considered.

$$\begin{pmatrix} (0,0) & (1,0) & (0,1) \\ (0,1) & (0,0) & (1,0) \\ (1,0) & (0,1) & (0,0) \end{pmatrix} \tag{5.7}$$

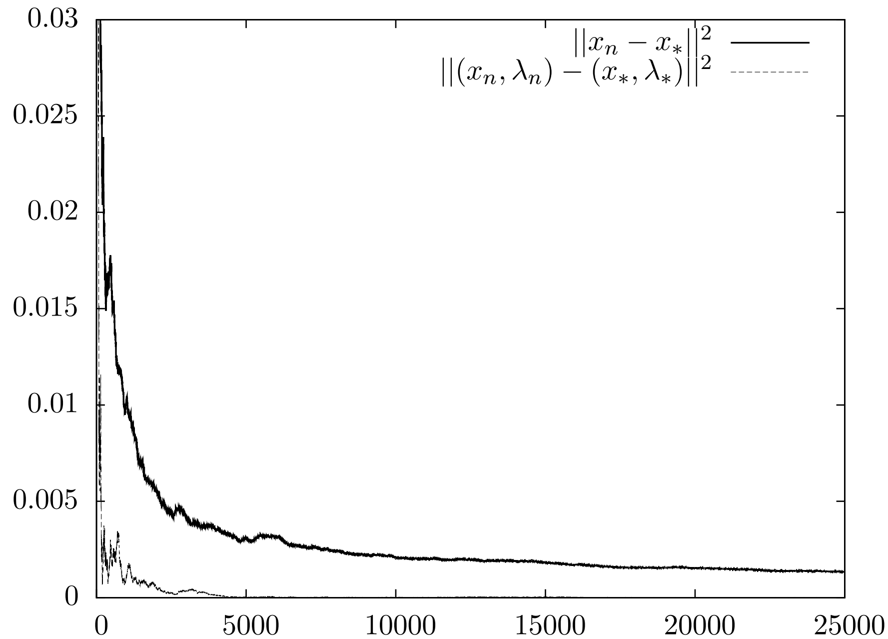

Figure 1: $\|(x_n, \lambda_n) - (x_*, \lambda_*)\|^2$ versus $\|x_n - x_*\|^2$.

5.2 Convergence with positive probability

The estimates given by Lemma 5.3 allows us to improve the range of parameters in which general convergence results can be obtained for the process (APD). We start by showing that there exists a unique rest point of (4.3) which is stable if $1 \leq 2\eta\alpha < 2$. Let $\mathcal{Y} \subseteq X \times \Delta$ be the set of rest points of (4.3) and let $B(\mathcal{A})$ be the basin of attraction corresponding to the attractor $\mathcal{A}$.

Proposition 5.8. *Assume that $1 \leq 2\eta\alpha < 2$. Then, there exists a unique rest point (x_*, λ_*) for the dynamics (4.3) which is an attractor.*

Proof. Let $(x_*, \lambda_*) \in \mathcal{Y}$. If $1 \leq 2\eta\alpha < 2$, equation (5.6) shows that the matrix $\nabla\Psi(x_*, \lambda_*)$ is stable. To prove that $\{(x_*, \lambda_*)\}$ is an attractor, take $V(x, \lambda) = ((x, \lambda) - (x_*, \lambda_*))^T D((x, \lambda) - (x_*, \lambda_*))$ as a (local) Lyapunov function where, for instance, D is the positive definite solution of the Lyapunov

equation $\nabla\Psi(x_*,\lambda_*)^T D + D\nabla\Psi(x_*,\lambda_*) = -I$. Given the fact that basins of attraction cannot overlap, the set $\mathcal{Y}$ is finite since $X\times\Delta$ is compact and Ψ is regular. Finally, $\mathcal{Y}$ reduces to one point since, in this case, it is impossible to have finitely many stable equilibibria due to the Poincaré–Hopf Theorem (see e.g., Milnor [20, Chapter 6]). ☐

The following definition is crucial to ensure convergence with positive probability of the process $(x_n,\lambda_n)_n$ to a given (not necessarily global) attractor.

Definition 5.9. *Let $(z_n)_n$ be a discrete stochastic process with state space Z. A point $z\in Z$ is attainable by $(z_n)_n$ if for each $m\in\mathbb{N}$ and every open neighborhood U of z, $\mathbb{P}(\exists n\geq m,\ z_n\in U)>0$.*

The following lemma uses strongly the particular form of the updating rule (AP) considered in this work.

Lemma 5.10. *Fix $\lambda=(\lambda^1,\ldots,\lambda^N)\in\Delta$. Set $x^i\in\mathbb{R}^{|S^i|}$ such that $x^{is}=G^i(s,\lambda^{-i})$ for all $s\in S^i$ and put $x=(x^1,\ldots,x^N)\in X$. Then, the point $(x,\lambda)\in X\times\Delta$ is attainable by the process $(x_n,\lambda_n)_n$. In particular, any rest point of the dynamics (4.3) is attainable.*

Proof. The fact that $\sigma_n^{is}\geq\xi>0$ for every $i\in A$, $s\in S^i$ and $n\in\mathbb{N}$ implies that any finite sequence generated by (APD) has positive probability. The updating rule (AP) can be expressed as (including only for this time the initial conditions)

$$x_{n+1}^{is}=\frac{1}{\theta_n^{is}+\theta_0^{is}}\big(g^i_{\upsilon^{is}(\theta_n^{is})}+g^i_{\upsilon^{is}(\theta_n^{is}-1)}+\cdots+g^i_{\upsilon^{is}(1)}+x_0^{is}\big),\tag{5.8}$$

where $\upsilon^{is}(k)=\inf\{q\geq 1,\ \theta_q^{is}=k\}$, i.e., the stage when player i has played strategy $s\in S^i$ for the k-th time. Observe that, without loss of generality, we can suppose that $m=0$ in the definition of attainability given the particular form of the updating rule (5.8).

Let $\zeta_n^{\mathbf{s}}$ be the number of times that the strategy profile $\mathbf{s}\in S$ has been played up to time n. Hence, for every $i\in A$ and $s\in S^i$, (5.8) implies that

$$\begin{aligned}x_{n+1}^{is}&=\sum_{\mathbf{r}\in S^{-i}}G^i(s,\mathbf{r})\frac{\zeta_n^{(s,\mathbf{r})}}{\theta_n^{is}+\theta_0^{is}}+\frac{x_0^{is}}{\theta_n^{is}+\theta_0^{is}},\\&=\sum_{\mathbf{r}\in S^{-i}}G^i(s,\mathbf{r})\frac{\zeta_n^{(s,\mathbf{r})}}{\theta_n^{is}}+\overline{b}_n,\end{aligned}$$

with $\overline{b}_n=O\big((\theta_n^{is})^{-1}\big)$. Observe that $\theta_n^{is}\to+\infty$ almost surely due to the conditional Borel–Cantelli lemma. Fix $\varepsilon>0$ and let n be an integer such that $k_s^i=n\tilde{k}_s^i\in\mathbb{N}$, where, for every $i\in A$ and $s\in S^i$, $\tilde{k}_s^i$ denotes a rational number satisfying that $|\lambda^{is}-\tilde{k}_s^i|<\varepsilon$. For a strategy profile $\mathbf{s}\in S$, let us define the positive integers $n_{\mathbf{s}}=\prod_{i\in A}k^i_{\mathbf{s}^i}$ and $\overline{n}=\sum_{\mathbf{s}\in S}n_{\mathbf{s}}$. Now we take the sequence generated by (APD) defined by $l\in\mathbb{N}$ blocks of size $\overline{n}$ where within each block, each $\mathbf{s}\in S$ is played exactly

$n_{\mathbf{s}}$ times, regardless the order of play. Fix $i \in A$ and $\mathbf{r} \in S^{-i}$, so that, by construction

$$
\begin{aligned}
\frac{\zeta_{l\overline{n}}^{(s,\mathbf{r})}}{\theta_{l\overline{n}}^{is}} &= \frac{\prod_{j\neq i} k_{\mathbf{r}^j}^{j} k_s^i}{k_s^i \sum\limits_{\mathbf{u}\in S^{-i}} \prod_{j\neq i} k_{\mathbf{u}^j}^{j}} \\
&= \frac{\prod_{j\neq i} \tilde{k}_{\mathbf{r}^j}^{j}}{\prod_{j\neq i} \big(\sum_{p_j=1}^{|S_j|} \tilde{k}_{p_j}^{j} \big)} \\
&= \prod_{j\neq i} \lambda^{j\mathbf{r}^j} + \tilde{b}_{\varepsilon},
\end{aligned}
$$

where $\tilde{b}_{\varepsilon} \to 0$ as $\varepsilon \to 0$. Finally, given $\varepsilon' > 0$, take l large and ε small to have $\|(x_{ln+1}, \lambda_{ln+1}) - (x, \lambda)\| < \varepsilon'$. ☐

Recall that $\mathcal{L}(z_n)$ is the limit set of the sequence $(z_n)_n$. The following result is the goal of this subsection.

Proposition 5.11. *If an attractor $\mathcal{A}$ for the dynamics (4.3) satisfies that $B(\mathcal{A}) \cap \mathcal{Y} \neq \emptyset$, then $\mathbb{P}(\mathcal{L}(x_n, \lambda_n) \subseteq \mathcal{A}) > 0$. In particular, under the Logit decision rule decision (5.1), if $1 \leq 2\eta\alpha < 2$, then $\mathcal{Y}$ reduces to one point (x_*, λ_*) and $\mathbb{P}((x_n, \lambda_n) \to (x_*, \lambda_*)) > 0$.*

Before providing the proof we need to briefly introduce the following concepts. Let ϕ be the semi-flow induced by the differential equation (4.3) and let Y_t the continuous time affine process associated to the discrete process $(x_n, \lambda_n)_n$, i.e.,

$$
Y(\tau_n + u) = (x_n, \lambda_n) + u \frac{(x_{n+1}, \lambda_{n+1}) - (x_n, \lambda_n)}{\tau_{n+1} - \tau_n}, \tag{5.9}
$$

for all $n \in \mathbb{N}$ and $u \in [0, \frac{1}{n+1})$, where $\tau_n = \sum_{m=1}^{n} \frac{1}{m}$. Let $(\mathcal{F}_t)_{t\geq 0}$ be the natural associated filtration.

The following technical lemma is needed. We only provide an outline of the proof because we follow exactly the lines in Benaïm [3, Proposition 4.1] along with the explicit computations provided in the proof of Schreiber [24, Theorem 2.6].

Lemma 5.12. *For all $T > 0$ and $\delta > 0$,*

$$
\mathbb{P}\bigg(\sup_{u\geq t} \big[\sup_{0\leq h\leq T} \|Y(u+h) - \phi_h(Y(u))\| \big] \geq \delta \mid \mathcal{F}_t \bigg) \leq \frac{C(\delta, T)}{\exp(ct)},
$$

for some positive constants c and $C(\delta, T)$ when $t \geq 0$ is large enough.

Outline of the proof. Roughly speaking, the process (APD) can be written as

$$
(x_{n+1}, \lambda_{n+1}) - (x_n, \lambda_n) = \frac{1}{n+1} \big(\Psi(x_n, \lambda_n) + \tilde{\epsilon}_{n+1} + b_{n+1} \big),
$$

where $(\epsilon_n)_n$ is almost surely bounded, $\mathbb{E}(\tilde{\epsilon}_{n+1} \mid \mathcal{F}_n) = 0$ and $b_n = O(1/n)$. Recall that $m(t)$ is the largest integer l such that $t \geq \sum\limits_{j=1}^{l} 1/j$. Then from Benaïm [3, Proposition 4.1] we have that

$$
\sup_{0\leq h\leq T} \|Y(u+h) - \phi_h(Y(u))\| \leq C(T) \bigg(\frac{1}{m(u)} + A(u, T) \bigg),
$$

for u sufficiently large, where

$$A(u,T) = \sup_{0\le h\le T} \left\| \sum_{l=m(u)}^{m(u+h)-1} \frac{\epsilon_{l+1}+b_{l+1}}{l+1} \right\|.$$

To conclude, it is enough to follow word by word, in this simpler case, the lines in the proof of Schreiber [24, Theorem 2.6]. □

Proof of Proposition 5.11. In view of Proposition 5.8 and Lemmas 5.10 and 5.12 the result follows directly from Benaïm [3, Theorem 7.3]. □

Remark 5.13. Notice that for Lemma 5.10 and for the first part of the statement in Proposition 5.11, we have only assumed the condition (A) on the decision rule σ. Furthermore, from Benaïm [3, Theorem 7.3], we have the following estimate for the probability of convergence to an attractor $\mathcal{A}$. If the set $U \subseteq X \times \Delta$ is such that $\overline{U} \subseteq B(\mathcal{A})$, then there exists numbers $T, \delta > 0$, depending on U so that

$$\mathbb{P}(\mathcal{L}(x_n,\lambda_n) \subseteq \mathcal{A}) \ge \left(1 - \frac{C(\delta,T)}{\exp(ct)}\right)\mathbb{P}(\exists\, u \ge t, Y(u) \in U),$$

for all $t > 0$, where the constants $C(\delta,T)$ and c are given by Lemma 5.12.

A traffic game

The (almost sure or with positive probability) convergence to attractors results obtained when the Logit decision rule is considered are valid under the strong assumption $2\eta\alpha < 2$. In fact, this condition becomes very difficult to verify as the number of players increases. Moreover, nonconvergence can occur for some games (see Section 5.3 for details) if the parameter $\eta\alpha$ is large. In this part, we will discuss the interesting application developed in Cominetti et al. [9, Section 3] and we will show that a result in the spirit of Proposition 5.11 can be obtained under a much weaker condition.

Consider a network with a topology that consists on a set of parallel routes. Each route $r \in \mathcal{R}$ in the network is characterized by an increasing sequence of values $c_1^r \le \cdots \le c_N^r$ where c_u^r represents the average travel time when r carries a load of u users.

The traffic game is defined as follows. The strategy set is common for all players, i.e., $S^i = \mathcal{R}$, for every $i \in A$ with $\mathcal{R}$ the set of available routes. The payoff to each player i, when the strategy profile $\mathbf{r} \in \mathcal{R}^N$ is played (i.e., when the network is loaded by the configuration $\mathbf{r}$), is given by the value $-c_u^{\mathbf{r}^i} = G^i(\mathbf{r})$, that is, minus her travel time.

This traffic game is shown to be a potential game in the sense that there exists a function $\Lambda : [0,1]^{N\times|\mathcal{R}|} \to \mathbb{R}$ such that

$$\frac{\partial \Lambda}{\partial \lambda^{is}}(\lambda) = G^i(s, \lambda^{-i}),$$

for every $\lambda \in \Delta$. Explicitly, the function Λ is given by

$$\Lambda(\pi) = -\mathbb{E}_\pi\big[\sum_{r\in\mathcal{R}} \sum_{u=1}^{U^r} c_u^r\big], \tag{5.10}$$

where the expectation is taken with respect to the random variables $U^r = \sum_{i \in A} X^{ir}$ with X^{ir} independent Bernouilli variables such that $\mathbb{P}(X^{ir} = 1) = \pi^{ir}$. It is also shown that the second derivatives of Λ are zero except for

$$\frac{\partial^2 \Lambda}{\partial \pi^{jr} \partial \pi^{ir}}(\pi) = \mathbb{E}_\pi \big(c^r_{U^r_{ij}+1} - c^r_{U^r_{ij}+2} \big) \in [-\eta, 0], \tag{5.11}$$

$i \neq j$, where $U^r_{ij} = \sum_{k \neq i,j} X^{kr}$.

Recall that we are assuming that player use the Logit decision rule in the learning process (APD). We also suppose that the smoothing parameters are identical for all players, i.e., $\beta_i = \beta$ for every $i \in A$. Notice that, in this framework, the value η (defined in (5.3)) translates to

$$\eta = \max\{\eta^r_u\,;\, r \in \mathcal{R},\ 2 \le u \le N\} = \max\{c^r_u - c^r_{u-1}\,;\, r \in \mathcal{R},\ 2 \le u \le N\}. \tag{5.12}$$

Cominetti et al. [9] obtain the following (among others) result. Recall that $\alpha = \max_{i \in A} \sum_{j \neq i} \beta_j = (N-1)\beta$.

Proposition 5.14. *The following hold*

(i) if $2\eta\alpha < 4$, the function F is contracting for the infinity norm and the process (4.4) converges almost surely to the unique rest point of (4.6),

(ii) if $\eta\beta < 1$, (4.6) has a unique rest point $x_* \in X$ which is symmetric in the sense that $x_* = (\hat{x}, \ldots, \hat{x})$. Furthermore, $\{x_*\}$ is an attractor for (4.6).

Notice that the first part in the proposition above can be recovered for the process (APD) since Proposition 5.2 applies. The second part provides a much weaker condition to have existence and uniqueness of a rest point of (4.6) (or, equivalently, a Nash equilibrium of the perturbed game defined in Lemma 5.1). Observe also that, despite the fact that the second part gives the existence of an attractor, no convergence result is obtained for the discrete process (4.4). The next result show that, under the assumption $\eta\beta < 1$, something more can be said for (APD).

Proposition 5.15. *If $\eta\beta < 1$, (4.3) has a unique rest point $(x_*, \lambda_*) \in X \times \Delta$ which is symmetric in the sense that $x_* = (\hat{x}, \ldots, \hat{x})$ and $\lambda_* = (\hat{\lambda}, \ldots, \hat{\lambda}) = \sigma(x_*)$. Furthermore, $\{(x_*, \lambda_*)\}$ is an attractor for (4.3) and $\mathbb{P}((x_n, \lambda_n) \to (x_*, \lambda_*)) > 0$.*

Proof. The existence and uniqueness of the symmetric rest point of (4.3) follows from Remark 4.1 and Proposition 5.14. Lemma 5.16 below shows that the matrix $\nabla\Psi(x_*, \lambda_*)$ is stable. Hence, $\{(x_*, \lambda_*)\}$ is an attractor for (4.3) and Proposition 5.11 applies. ☐

Lemma 5.16. *If $\eta\beta < 1$, then the matrix $\nabla\Psi(x_*, \lambda_*)$ is stable.*

Proof. Recall that the matrix $J^\beta = \nabla_x \Psi_x(x_*, \lambda_*)$ is the upper-left block of the matrix $\nabla\Psi(x_*, \lambda_*)$ (see (5.5)). Observe that, from the definition of the function Ψ_x, the fact that σ^i depends only on

x^i and (5.10), the entries of J^β are given by

$$
\begin{aligned}
J^\beta_{is,jr} = \frac{\partial \Psi^{is}_x}{\partial x^{jr}}(x_*, \lambda_*) &= \frac{\partial}{\partial x^{jr}}\left(\frac{\partial \Lambda}{\partial \pi^{is}}(\sigma(x))\right)(x_*) - \mathbb{1}_{\{is=jr\}} \\
&= \sum_{k\in A}\sum_{r'\in\mathcal{R}} \frac{\partial^2 \Lambda}{\partial \pi^{kr'}\partial \pi^{is}}(\lambda_*)\frac{\partial \sigma^{kr'}}{\partial x^{jr}}(x_*) - \mathbb{1}_{\{is=jr\}} \\
&= \sum_{r'\in\mathcal{R}} \frac{\partial^2 \Lambda}{\partial \pi^{jr'}\partial \pi^{is}}(\lambda_*)\frac{\partial \sigma^{jr'}}{\partial x^{jr}}(x_*) - \mathbb{1}_{\{is=jr\}} \\
&= \frac{\partial^2 \Lambda}{\partial \pi^{jr}\partial \pi^{ir}}(\lambda_*)\frac{\partial \sigma^{jr}}{\partial x^{jr}}(x_*)\mathbb{1}_{\{s=r,i\neq j\}} - \mathbb{1}_{\{is=jr\}} \\
&= \beta\lambda^{jr}_*(1-\lambda^{jr}_*)\mathbb{E}_{\lambda_*}\big(c^r_{U^r_{ij}+1} - c^r_{U^r_{ij}+2}\big)\mathbb{1}_{\{s=r,i\neq j\}} - \mathbb{1}_{\{is=jr\}}.
\end{aligned} \tag{5.13}
$$

Since λ_* is symmetric ($\lambda^{ir} = \lambda^{jr}$, for all $i,j\in A$), J^β is a symmetric matrix. Let us show that J^β is negative definite by modifying the trick used in Cominetti et al. [9, Proposition 12]. Take $h\in\mathbb{R}^{N|\mathcal{R}|}\backslash\{0\}$, then, from (5.13),

$$
h^T J^\beta h = \sum_{r\in\mathcal{R}}\big[\beta\sum_{i\neq j} h^{ir}\sqrt{\lambda^{ir}_*(1-\lambda^{ir}_*)}h^{jr}\sqrt{\lambda^{jr}_*(1-\lambda^{jr}_*)}\mathbb{E}_{\lambda_*}\big(c^r_{U^r_{ij}+1} - c^r_{U^r_{ij}+2}\big) - \sum_i (h^{ir})^2\big].
$$

For every $i\in A$ and $r\in\mathcal{R}$, put $v^{ir} = h^{ir}\sqrt{\frac{1-\lambda^{ir}_*}{\lambda^{ir}_*}}$, $Z^{ir} = v^{ir}X^{ir}$ and set $\eta^r_0 = \eta^r_1 = 0$. Therefore,

$$
\begin{aligned}
h^T J^\beta h &= \sum_{r\in\mathcal{R}}\left[\beta\sum_{i\neq j} v^{ir}v^{jr}\lambda^{ir}_*\lambda^{jr}_*\mathbb{E}_{\lambda_*}\big(c^r_{U^r_{ij}+1} - c^r_{U^r_{ij}+2}\big) - \sum_i \lambda^{ir}_*\frac{(v^{ir})^2}{1-\lambda^{ir}_*}\right] \\
&= \sum_{r\in\mathcal{R}}\mathbb{E}_{\lambda_*}\left(\beta\sum_{i\neq j} Z^{ir}Z^{jr}(c^r_{U^r-1} - c^r_{U^r}) - \sum_i \frac{(Z^{ir})^2}{1-\lambda^{ir}_*}\right) \\
&\leq \sum_{r\in\mathcal{R}}\mathbb{E}_{\lambda_*}\Big(-\eta^r_{U^r}\beta\sum_{i\neq j} Z^{ir}Z^{jr} - \sum_i (Z^{ir})^2\Big) \\
&= \sum_{r\in\mathcal{R}}\mathbb{E}_{\lambda_*}\left(-\eta^r_{U^r}\beta\Big(\sum_i Z^{ir}\Big)^2 + (\eta^r_{U^r}\beta - 1)\sum_i (Z^{ir})^2\right) < 0,
\end{aligned}
$$

where the last inequality follows by observing that $\eta^r_{U^r}\leq\eta$. ☐

Remark 5.17. In fact, part (ii) in Proposition 5.14 holds true if $\eta\beta < 2$. The authors provide an explicit (local) Lyapunov function which does not seem suitable in our case.

5.3 Nonconvergence

In order to give an idea of the behavior of the stochastic process defined by (APD) when the parameter β (we assume $\beta_i = \beta$ for all $i\in A$) becomes large, we provide a small class of games which underlines the relevance of the hypotheses considered throughout this document. Consider a 2-player symmetric game, i.e., the strategy set $\overline{S} = S^1 = S^2$ is common for and the payoffs verify that $G^1 = (G^2)^T$. Let us assume that G^1 has constant-sum by row, this is, $\sum_r G^1(s,r) = k\in\mathbb{R}$ for every $s\in\overline{S}$. It is easy to check that for this kind of games there exists a rest point of (4.3) which

has the form $(\overline{x}, \sigma(\overline{x})) \in X \times \Delta$ such that $\overline{x}^i = (1/k, \dots, 1/k)$ for $i \in \{1, 2\}$. We also assume that $\sum_s G^1(s, s) \neq k$.

A game that satisfies the precedent conditions is the *good* (resp. *bad*) Rock-Scissors-Paper game

$$\begin{pmatrix} 0 & a & -b \\ -b & 0 & a \\ a & -b & 0 \end{pmatrix},$$

where $0 < b < a$ (resp. $0 < a < b$) or the game (5.7).

The (strong) hypotheses above ensure that at least one rest point of (4.3) does not depend on the parameter β. In the following we will easily show that if the parameter β is sufficiently large then the rest point $(\overline{x}, \sigma(\overline{x}))$ becomes linearly unstable. Later, we will prove that this implies that $\mathbb{P}((x_n, \lambda_n) \to (\overline{x}, \sigma(\overline{x}))) = 0$.

Lemma 5.18. *If the parameter $\beta > 0$ is sufficiently large, then there exists an eigenvalue μ of the matrix $\nabla\Psi(\overline{x}, \sigma(\overline{x}))$ such that* $\mathrm{Re}(\mu) > 0$.

Proof. Again, let $J^\beta = \nabla_x \Psi_x(\overline{x}, \sigma(\overline{x}))$ be the upper-left block of the Jacobian matrix of the function Ψ, which is the only relevant part, evaluated at $(\overline{x}, \sigma(\overline{x}))$. The precise expression for the entries of J^β is

$$J^\beta_{is,jr} = \frac{\partial \Psi^{is}_x}{\partial x^{jr}}(\overline{x}, \sigma(\overline{x})) = \begin{cases} -1, & \text{if } i = j \text{ and } s = r \\ \;\; 0, & \text{if } i = j \text{ and } s \neq r \\ \beta \frac{1}{|\overline{S}|}\left[G^i(s, r) - \frac{k}{|\overline{S}|}\right], & \text{otherwise,} \end{cases}$$

$i, j \in \{1, 2\}$. Thus the matrix J^β has the form

$$\begin{pmatrix} -I & \overline{J}^\beta \\ \overline{J}^\beta & -I \end{pmatrix},$$

with $\overline{J}^\beta \in \mathbb{R}^{|\overline{S}|} \times \mathbb{R}^{|\overline{S}|}$. Observe that we can decompose J^β as $J^\beta = \beta J - I$, where

$$J = \begin{pmatrix} \mathbf{0} & \overline{J} \\ \overline{J} & \mathbf{0} \end{pmatrix}.$$

Let $\mu_1, \dots, \mu_{|S|} \in \mathbb{C}$ be the eigenvalues of $\overline{J}$ (counting multiplicity). Since we have assumed that $\sum_s G^1(s, s) \neq k$, the trace of $\overline{J}$ is not zero. So that, there exists some eigenvalue μ_k, $k \in \{1, \dots |\overline{S}|\}$, with nonzero real part. We have that, if v is an eigenvector associated to μ_k, then μ_k is an eigenvalue of J with corresponding eigenvector $\mathbf{u} = (v, v) \in \mathbb{R}^{|\overline{S}|} \times \mathbb{R}^{|\overline{S}|}$ since

$$J\mathbf{u} = \begin{pmatrix} \mathbf{0} & \overline{J} \\ \overline{J} & \mathbf{0} \end{pmatrix} \begin{pmatrix} v \\ v \end{pmatrix} = \begin{pmatrix} \overline{J}v \\ \overline{J}v \end{pmatrix} = \mu_k \mathbf{u}.$$

If $\mathrm{Re}(\mu_k) > 0$, the proof is finished. If $\mathrm{Re}(\mu_l) \leq 0$ for all $l \in \{1, \dots |\overline{S}|\}$ then $\sum_l \mathrm{Re}(\mu_l) < 0$. Also, the trace of J is zero and therefore there exists μ eigenvalue of J (which is not an eigenvalue of $\overline{J}$) such that $\mathrm{Re}(\mu) > 0$.

Finally, observe that

$$\det(\beta J - \mu I) = \frac{1}{\beta^{|\overline{S}|}} \det\left(J - \frac{\mu}{\beta} I\right), \tag{5.14}$$

and it is straightforward from (5.14) that $\overline{\mu}$ is an eigenvalue of the matrix J^{β} if $\mu = (1+\overline{\mu})/\beta$ is an eigenvalue of $\overline{J}$. Then $\overline{\mu} = \beta\mu - 1$ whose real part is strictly positive for a sufficiently large β. □

Proposition 5.19. *There exists $\beta > 0$ large enough and at least one rest point $(\overline{x}, \sigma(\overline{x})) \in X \times \Delta$ of* (4.3) *such that,*

$$\mathbb{P}((x_n, \lambda_n) \to (\overline{x}, \sigma(\overline{x}))) = 0.$$

Proof. We can directly apply Brandière and Duflo [6, Theorem 1]. The hypotheses of the theorem concerning the continuous dynamics and the step-size of the discrete process (APD) are immediately satisfied. The only condition that deserves attention is the one referred to the excitation of the noise on a repulsive direction at $(\overline{x}, \sigma(\overline{x}))$. Explicitly, it is sufficient to prove that

$$\liminf_{n\to+\infty} \mathbb{E}(||\epsilon^{pr}_{n+1}||^2 \mid \mathcal{F}_n) > 0 \text{ a.s. on the event } \Gamma\big(\overline{x}, \sigma(\overline{x})\big) = \{(x_n, \lambda_n) \to (\overline{x}, \sigma(\overline{x}))\}, \tag{5.15}$$

since the noise term $\epsilon_n = (U_n, M_n)$ is almost surely bounded. Here the upper-script pr stands for the projection onto the repulsive subspace spanned by the eigenvectors associated to the eigenvalues with positive real part.

Fix $i \in \{1, 2\}$, take β large to have an eigenvalue μ of $\nabla\Psi(\overline{x}, \sigma(\overline{x}))$ such that $\mathrm{Re}(\mu) > 0$ and let $\mathbf{v}$ a correspondent (eventually generalized) eigenvector. The vector $\mathbf{v}$ has the form $\mathbf{v} = (v_1, v_2)$. Notice that necessarily $v_2 \neq 0$ since, if $v_2 = 0$, then v_1 is a vector of ones, which is indeed an eigenvector for the upper-left block of $\nabla\Psi(\overline{x}, \sigma(\overline{x}))$ having -1 as the associated eigenvalue. So that

$$\begin{aligned}
\mathbb{E}(||\varepsilon^{pr}_{n+1}||^2 \mid \mathcal{F}_n) &\geq \mathbb{E}(\|\langle \varepsilon_{n+1}, \mathbf{v}\rangle \mathbf{v}\|^2 \mid \mathcal{F}_n) \\
&\geq \mathbb{E}(\sum_r (v_2^{jr} M^{jr}_{n+1})^2 \mid \mathcal{F}_n) \\
&\geq c\mathbb{E}((M^{jr}_{n+1})^2 \mid \mathcal{F}_n),
\end{aligned}$$

with $j = -i$ and for some $r \in \overline{S}$ and $c > 0$.

In view of (4.2),

$$\mathbb{E}((M^{jr}_{n+1})^2 \mid \mathcal{F}_n) = \mathbb{E}((\mathbb{1}_{\{s^j_{n+1}=r\}} - \sigma^{jr}(x^j_n))^2 \mid \mathcal{F}_n) + O\left(\frac{1}{n}\right) = \sigma^{jr}(x^j_n)(1 - \sigma^{jr}(x^j_n)) + O\left(\frac{1}{n}\right).$$

Finally, take the $\liminf_n$ in the previous expression on the event $\Gamma(\overline{x}, \sigma(\overline{x}))$ to conclude that (5.15) holds, since σ^{is} is bounded away from zero for every $i \in \{1, 2\}$ and $s \in \overline{S}$. □

Remark 5.20. For the game (5.7), if $\beta > 3/2$, then $\rho(\nabla\Psi(\overline{x}, \sigma(\overline{x}))) > 0$. Compare with Propositions 5.2 and 5.11.

As observed by Pemantle [21], the nonconvergence results like the previous proposition are not very interesting if the set of unstable points is *too large*. Particularly, the most useful consequences can be stated when this set is finite. This is the case of our example (5.3) and, moreover, it is easy to check that $(\overline{x}, \sigma(\overline{x}))$ is the unique rest point of (4.3) for all $\beta > 0$. The previous result shows that for a large β, $(\overline{x}, \sigma(\overline{x}))$ has probability zero to be the limit of the process while for small β it is almost surely the limit. Simulations suggest that there is a cycle that attracts the trajectories and that the empirical frequencies of play still converge to $\sigma(\overline{x})$. Figure 2 shows the behavior of the procedure (specifically the evolution of the mixed action σ^1_n of player 1) when β is large.

However, it does not seem plausible to pursuit a result like Proposition 5.19 on a general class of games. For instance, consider the 2-player zero-sum game defined by the payoff

$$G = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{5.16}$$

Let $(x_*, \sigma(x_*))$, with $\sigma^1(x_*^1) = \sigma^2(x_*^2) = (1/(1+e^\beta), e^\beta/(1+e^\beta))$ and $x_*^1 = x_*^2 = (-e^\beta/(1+e^\beta), 1/(1+e^\beta))$, be the unique rest point of (4.3). In this case, every eigenvalue of $\nabla\Psi(x_*, \sigma(x_*))$ is equal to -1. Then $\mathbb{P}((x_n, \lambda_n) \to (x_*, \sigma(x_*))) > 0$ for all $\beta > 0$ due to Proposition 5.11.

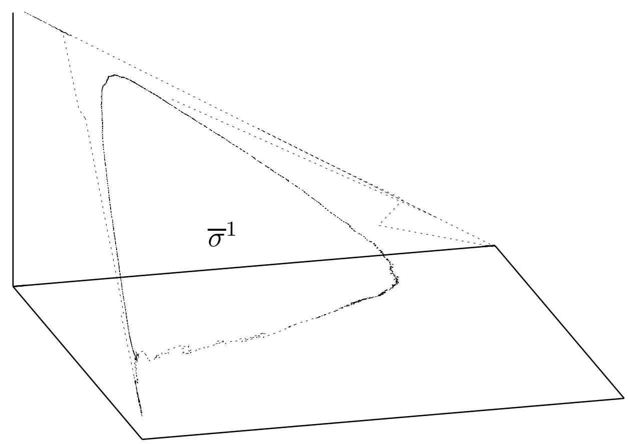

Figure 2: The mixed action σ_n^1 of Player 1 when $\beta = 4$.

6 Random environment

Our aim in this section is to consider the case where players receive, at each stage, a perturbed version of their payoffs. The general result given by Theorem 2.2 allows to add some perturbation to the process (APD) maintaining unaltered the results presented in this work. For example, we can consider that each player $i \in A$ get a payoff $\overline{g}_n^i = g_n^i + \varepsilon_n^i$ at stage n, where $(\varepsilon_n^i)_n$ is a martingale difference process bounded in L^2 or a vanishing random variable (see Remark 2.3). In this section, we are interested in a different kind of perturbation.

For the sequel, the model runs analogously as before. We only add that, at each time $n \in \mathbb{N}$, each player $i \in A$ receives a random payoff $\overline{g}_n^i = G^i(s_n, w_n)$ where the sequence $(w_n)_n$ is a controlled (by the parameter $\lambda \in \Delta$) Markov chain with finite state space W, i.e., there exists a family of transition matrices $(P(\lambda))_{\lambda\in\Delta}$ where

$$\mathbb{P}(w_{n+1} = w \mid \mathcal{F}_n) = P_{(w_n,w)}(\lambda_n),$$

for $w \in W$ with λ_n determined by the learning process and $\mathcal{F}_n$ being the σ-algebra generated by the history $(\mathbf{s}_1, g_1, w_1, \ldots, \mathbf{s}_n, g_n, w_n)$ up to time n. We assume the following.

(A1) For any $\lambda \in \Delta$ the Markov chain with transition matrix $P(\lambda)$ has a unique invariant probability $\tau(\lambda) \in \Delta^W$, where Δ^W denotes the set of probabilities over W.

(A2) The function $P : \Delta \to \mathcal{M}$ (the set of stochastic matrices of dimension $|W| \times |W|$), $\lambda \to P(\lambda)$ is of class $\mathcal{C}^1$.

Remark 6.1. The hypothesis above imply that the function τ is also of class $\mathcal{C}^1$ (see Benaïm [2, Lemma 3.2]).

Observe that the unique recurrent class of the associated Markov chain may be periodic. Note also that the process $(\mathbf{s}_n, w_n)_n$ is also a controlled (by the parameter $(x, \lambda) \in X \times \Delta$) Markov chain with state space $S \times W$. The independence hypothesis implies that for a given (x, λ) the associated transition matrix is given by

$$\overline{P}_{(\mathbf{a},\mathbf{b})}(x, \lambda) = \prod_{i \in A} \sigma^{i\mathbf{r}^i}(x^i) P_{(w,w')}(\lambda),$$

for each $(\mathbf{a}, \mathbf{b}) = \big((\mathbf{s}, w), (\mathbf{r}, w')\big) \in (S \times W)^2$. From (A1), the Markov chain with transition matrix $\overline{P}(x, \lambda)$ has a unique invariant probability $\overline{\tau}(x, \lambda)$, where

$$\overline{\tau}_{\mathbf{a}}(x, \lambda) = \prod_{i=1}^{N} \sigma^{i\mathbf{s}^i}(x^i) \tau_w(\lambda), \tag{6.1}$$

for $\mathbf{a} = (\mathbf{s}, w) \in S \times W$.

Let us now state precisely the corresponding discrete process in this framework. We conserve the notation and the hypotheses (A) over the decision rule of players σ. For each player $i \in A$, we define the new payoff function $G^i : S \times W \to \mathbb{R}$ and its multilinear extension to $\Delta \times \Delta^W$ as usual. Therefore the updating rule is this case is given by

$$x^{is}_{n+1} = \begin{cases} (1 - \frac{1}{\theta^{is}_n}) x^{is}_n + \frac{1}{\theta^{is}_n} \overline{g}^i_{n+1}, & \text{if } s = s^i_{n+1}, \\ x^{is}_n, & \text{otherwise}, \end{cases} \tag{APM}$$

with $\overline{g}^i_n = G^i(\mathbf{s}_n, w_n)$. As before, we can conveniently recast the process (APM) as

$$\begin{cases} x^{is}_{n+1} - x^{is}_n = \dfrac{1}{n+1} \left[\dfrac{\sigma^{is}(x^i_n) G^i(s, \sigma^{-i}(x_n), \tau(\lambda_n))}{\lambda^{is}_n} + \dfrac{\overline{U}^{is}_{n+1}}{\lambda^{is}_n} \right], \\ \\ \lambda^{is}_{n+1} - \lambda^{is}_n = \dfrac{1}{n+1} \left[\sigma^{is}(x^i_n) - \lambda^{is}_n + M^{is}_{n+1} \right], \end{cases} \tag{APDM}$$

where

$$\begin{aligned} \overline{U}^{is}_{n+1} &= (\overline{g}^i_{n+1} - x^{is}_n) \mathbb{1}_{\{s = s^i_{n+1}\}} - \big(\sigma^{is}(x^i_n) G^i(s, \sigma^{-i}(x_n), \tau(\lambda_n)) - x^{is}_n\big), \\ M^{is}_{n+1} &= \mathbb{1}_{\{s = s^i_{n+1}\}} - \sigma^{is}(x^i_n) + b_{n+1}, \end{aligned} \tag{6.2}$$

and $b_n = O(\frac{1}{n})$. Observe that the process $(\overline{U}_n)_n$ is not a martingale difference sequence. The next result shows that the analogous to Proposition 4.5 holds.

Proposition 6.2. *The process* (APDM) *converges almost surely to an* ICT *set of the continuous dynamics*

$$\begin{cases} \dot{x}_t^{is} = \dfrac{\sigma^{is}(x_t^i)}{\lambda_t^{is}}\Big[G^i(s,\sigma^{-i}(x_t),\tau(\lambda)) - x_t^{is}\Big] & = \overline{\Psi}_x^{is}(x,\lambda), \\ \\ \dot{\lambda}_t^{is} = \sigma^{is}(x_t^i) - \lambda_t^{is} & = \overline{\Psi}_\lambda^{is}(x,\lambda). \end{cases} \tag{6.3}$$

In the sequel, our analysis will rely on the ideas in the proof of Benaïm [2, Proposition 3.3] (which are indeed a reduction to the finite state space case of the general framework developed in Benveniste et al. [4, Part II]) to show that hypothesis (c) of Theorem 2.2 holds (since assumptions (a) and (b) are immediately satisfied).

Fix $i \in \{1,\dots,N\}$ and $(x,\lambda) \in X \times \Delta$. Let us define the matrices $\mathbf{H}^i(x,\lambda)$ and $\mathbf{W}(x,\lambda)$ by

$$\mathbf{H}^i_{(\mathbf{a},\mathbf{b})}(x,\lambda) = \begin{cases} G^i(\mathbf{s},w) - x^{i\mathbf{s}^i}, & \text{if } \mathbf{s}^i = \mathbf{r}^i, \\ 0, & \text{otherwise}, \end{cases} \tag{6.4}$$

and

$$\mathbf{W}_{(\mathbf{a},\mathbf{b})}(x,\lambda) = \overline{\tau}_{\mathbf{b}}(x,\lambda), \tag{6.5}$$

for every $\mathbf{a} = (\mathbf{s},w), \mathbf{b} = (\mathbf{r},w') \in S \times W$ and $\overline{\tau}_{\mathbf{b}}(x,\lambda)$ given by equation (6.1). Notice that, for every $s \in S^i$, $\mathbf{a}_{n+1} = (\mathbf{s}_{n+1}, w_{n+1})$ and $\mathbf{b}_{n+1} = \big((s,\mathbf{s}_{n+1}^{-i}), w_{n+1}\big)$, the following equality holds

$$\mathbf{H}^i_{(\mathbf{a}_{n+1},\mathbf{b}_{n+1})}(x_n,\lambda_n) = \big(G^i(s,\mathbf{s}_{n+1}^{-i},w_{n+1}) - x_n^{is}\big)\mathbb{1}_{\{s=\mathbf{s}_{n+1}^i\}},$$

and

$$\begin{aligned} \big[\mathbf{W}(x_n,\lambda_n)\mathbf{H}^i(x_n,\lambda_n)\big]_{(\mathbf{a}_{n+1},\mathbf{b}_{n+1})} &= \sum_{\mathbf{a}=(\mathbf{r},w)} \overline{\tau}_{\mathbf{a}}(x_n,\lambda_n)\big(G^i(\mathbf{r},w) - x_n^{i\mathbf{r}^i}\big)\mathbb{1}_{\{s=\mathbf{r}^i\}} \\ &= \sigma^{is}(x_n^i)\Big(\sum_w \tau_w(\lambda_n) G^i(s,\sigma^{-i}(x_n),w) - x_n^{is}\Big) \\ &= \sigma^{is}(x_n^i)\Big(G^i(s,\sigma^{-i}(x_n),\tau(\lambda_n)) - x_n^{is}\Big), \end{aligned}$$

so that

$$\big[(I - \mathbf{W}(x_n,\lambda_n))\mathbf{H}^i(x_n,\lambda_n)\big]_{(\mathbf{a}_{n+1},\mathbf{b}_{n+1})} = \overline{U}_{n+1}^{is}. \tag{6.6}$$

Lemma 6.3. *There exists a $\mathcal{C}^1$ function $Q^i(x,\lambda)$ such that*

$$(I - \overline{P}(x,\lambda))Q^i(x,\lambda) = (I - \mathbf{W}(x,\lambda))\mathbf{H}^i(x,\lambda). \tag{6.7}$$

Outline of the proof. The required function is given by $Q^i(x,\lambda) = E(x,\lambda)\mathbf{H}^i(x,\lambda)$ where

$$E(x,\lambda) = \int_0^{+\infty} (E_t(x,\lambda) - \mathbf{W}(x,\lambda))dt,$$

with $E_t(x,\lambda)$ being the matrix solution of the linear differential equation

$$\begin{aligned} \frac{d}{dt}E_t^i(y) &= -\,(I - \overline{P}(y))E_t^i(y), \\ E_0^i(y) &= I. \end{aligned}$$

See Benaïm [2, Lemma 5.1] (and the previous discussion therein) for further details. □

Proof of Proposition 6.2. To prove that the third assumption in the statement of Theorem 2.2 is satisfied almost surely, we have to show that the noise term $(\epsilon_n)_n = (U_n, M_n)_n$, where $U_n^{is} = \overline{U}_n^{is}/\lambda_n^{is}$ (see equation (6.2)) satisfies that

$$\sup\left\{ \left\| \sum_{l=n}^{k-1} \frac{1}{l+1}\epsilon_{l+1} \right\|; k \in \{n+1, \dots, m(\sum_{j=1}^{n} \frac{1}{j} + T)\} \right\} \longrightarrow 0, \tag{6.8}$$

as $n \to +\infty$, for every $T > 0$. Since $(M_n)_n$ is an almost surely bounded martingale difference process plus a vanishing term, it is sufficient to show that (6.8) holds for $(U_n)_n$. For a matrix $\mathcal{B}$ (of appropriate dimension) and a fixed $s \in S^i$, let us set by $\mathcal{B}[n]$ the operation that consists on taking the $(\mathbf{a}_n, \mathbf{b}_n)$ entry of $\mathcal{B}$, where $\mathbf{a}_n = (\mathbf{s}_n, w_n)$ and $\mathbf{b}_n = \big((s, \mathbf{s}_n^{-i}), w_n\big)$. Now, according to (6.6) and Lemma 6.3, we have, for every $n \geq 1$,

$$\begin{aligned}
\frac{U_{n+1}^{is}}{n+1} =& \frac{1}{(n+1)\lambda_n^{is}}\big((I - \mathbf{W}(x_n, \lambda_n))\mathbf{H}^i(x_n, \lambda_n)\big)[n+1] \\
=& \frac{1}{(n+1)\lambda_n^{is}} Q^i(x_n, \lambda_n)[n+1] - \frac{1}{(n+1)\lambda_n^{is}} \overline{P}(x_n, \lambda_n) Q^i(x_n, \lambda_n)[n+1] \\
=& u_{n+1}^{is}(1) + u_{n+1}^{is}(2) + u_{n+1}^{is}(3) + u_{n+1}^{is}(4),
\end{aligned}$$

where

$$\begin{aligned}
u_{n+1}^{is}(1) &= \frac{1}{(n+1)\lambda_n^{is}} Q^i(x_n, \lambda_n)[n+1] - \frac{1}{(n+1)\lambda_n^{is}} \overline{P}(x_n, \lambda_n) Q^i(x_n, \lambda_n)[n], \\
u_{n+1}^{is}(2) &= \frac{1}{(n+1)\lambda_n^{is}} \overline{P}(x_n, \lambda_n) Q^i(x_n, \lambda_n)[n] - \frac{1}{n\lambda_{n-1}^{is}} \overline{P}(x_n, \lambda_n) Q^i(x_n, \lambda_n)[n], \\
u_{n+1}^{is}(3) &= \frac{1}{n\lambda_{n-1}^{is}} \overline{P}(x_n, \lambda_n) Q^i(x_n, \lambda_n)[n] - \frac{1}{(n+1)\lambda_n^{is}} \overline{P}(x_{n+1}, \lambda_{n+1}) Q^i(x_{n+1}, \lambda_{n+1})[n+1], \quad \text{and} \\
u_{n+1}^{is}(4) &= \frac{1}{(n+1)\lambda_n^{is}} \overline{P}(x_{n+1}, \lambda_{n+1}) Q^i(x_{n+1}, \lambda_{n+1})[n+1] + \\
&\qquad - \frac{1}{(n+1)\lambda_n^{is}} \overline{P}(x_n, \lambda_n) Q^i(x_n, \lambda_n)[n+1].
\end{aligned}$$

Note that, almost surely, $\left\|u_{n+1}^{is}(2)\right\| \leq C/n^2$ and $\left\|\sum_{l=n}^{k-1} u_{l+1}^{is}(3)\right\| \leq C/n$ since the functions $\overline{P}, Q^i$, and the process $(\lambda_n)_n$ are bounded (for a generic positive constant C). We also have that $\left\|u_{n+1}^{is}(4)\right\| \leq \|(x_{n+1}, \lambda_{n+1}) - x_n, \lambda_n\| \cdot C/n \leq C/n^2$ because of the smoothness of $\overline{P}Q^i$ and the compactness of $X \times \Delta$ and that $(\sum_{l=1}^{n} u_{l+1}^{is}(1))_n$ is a convergent martingale (since $\left\|u_{l+1}^{is}(1)\right\|^2 \leq C/l^2$). So that

$$\left\| \sum_{l=n}^{k-1} \frac{1}{l+1} U_{l+1}^{is} \right\| = \left\| \sum_{l=n}^{k-1} u_{l+1}^{is}(1) \right\| + O\left(\frac{1}{n}\right),$$

which implies that (6.8) holds. □

Once we have the connection between the discrete process (APDM) and the continuous dynamics (6.3), the task of finding explicit almost sure convergence results becomes more difficult because of the general form of the function P defined in (A2). However, it is possible to state an analog of Proposition 4.5 in the Logit decision rule case by strengthening the conditions over the parameters β_i and assuming that the variation of payoffs is small. Recall that $\alpha = \max_{i \in A} \sum_{j \neq i} \beta_j$.

Proposition 6.4. *Under the Logit decision rule* (5.1)*, let η_w be the maximal unilateral deviation that a player can face for a fixed $w \in W$ (see* (5.3)*). Set $\overline{\eta} = \max_w \eta_w$ and $\tilde{\eta} = \max_{i,\mathbf{s},w,w'} |G^i(\mathbf{s},w) - G^i(\mathbf{s},w')|$. If*

$$\begin{cases} 2\overline{\eta}\alpha + \tilde{\eta}\sqrt{|W|}k < 1, \\ 2\max_i \beta_i < 1, \end{cases} \tag{6.9}$$

where k is the $\|\cdot\|_\infty$-Lipschitz constant for the function τ, then the discrete process (APDM) *converges almost surely to the global attractor $\{(x_*, \lambda_*)\}$ for the dynamics* (6.3)*.*

Proof. It is not difficult to see that under condition (6.9) there exists a unique rest point $(x_*, \lambda_*) \in X \times \Delta$ of (6.3). Consider the function V defined by $V(x, \lambda) = \max\{\|x - x_*\|_\infty, \|\lambda - \lambda_*\|_\infty\}$. Now if $V(x_t, \lambda_t) = \|\lambda_t - \lambda_*\|_\infty \geq \|x_t - x_*\|_\infty$ then $\dot{V}(x_t, \lambda_t) \leq -KV(x_t, \lambda_t)$ for some $K \geq 0$, using exactly the same computations as in Proposition 4.5,the fact that $2\max_i \beta_i < 1$ and that $\left\|\sigma^{is}(x) - \sigma^{is}(y)\right\|_\infty \leq 2\beta_i \|x - y\|_\infty$ for all $i \in A$ and $s \in S^i$ (see e.g., Cominetti et al. [9, Proposition 5]). On the other hand, if $V(x_t, \lambda_t) = \|x_t - x_*\|_\infty \geq \|\lambda - \lambda_*\|_\infty$, assume that the maximum is attained on the is coordinate, that $x^{is} \geq x_*^{is}$ and define $F^{is}(x, \tau) = G^i(s, \sigma^{-i}(x), \tau)$. So that

$$\begin{aligned} \frac{d}{dt}(x_t^{is} - x_*^{is}) &= \frac{\sigma^{is}(x_t^i)}{\lambda_t^{is}}\Big[x_*^{is} - x_t^{is} + F^{is}(x_t, \tau(\lambda_t)) - F^{is}(x_*, \tau(\lambda_t)) + \\ &\qquad + F^{is}(x_*, \tau(\lambda_t)) - F^{is}(x_*, \tau(\lambda_*))\Big] \\ &\leq \frac{\sigma^{is}(x_t^i)}{\lambda_t^{is}}\Big[- \|x_t - x_*\|_\infty + 2\overline{\eta}\alpha \|x_t - x_*\|_\infty + \tilde{\eta}\, \|\tau(\lambda_t) - \tau(\lambda_*)\|_1 \Big] \\ &\leq -\xi\big(1 - 2\overline{\eta}\alpha - \tilde{\eta}\sqrt{|W|}k\big)V(x_t, \lambda_t), \end{aligned}$$

for some $\xi > 0$ such that $\sigma^{is}(x) \geq \xi$ for every $i \in A$ and $s \in S^i$. Notice that we have used the fact that, for every $w' \in W$,

$$\begin{aligned} F^{is}(x_*, \tau(\lambda_t)) - F^{is}(x_*, \tau(\lambda_*)) &= \sum_{w \in W} (\tau_w(\lambda_t) - \tau_w(\lambda_*))(F^{is}(x_*, w) - F^{is}(x_*, w')), \\ &\leq \tilde{\eta}\, \|\tau(\lambda_t) - \tau(\lambda_*)\|_1 \\ &\leq \tilde{\eta}\sqrt{|W|}\, \|\tau(\lambda_t) - \tau(\lambda_*)\|_\infty \leq \tilde{\eta}\sqrt{|W|}k\, \|\lambda_t - \lambda_*\|_\infty . \end{aligned}$$

Again, if $x_t^{is} - x_*^{is} < 0$, the computations are the same. Then, V is a strict Lyapunov function. □

The constant case

We restrict our attention to the constant case, i.e., the function P is such that $P(\lambda) \equiv P^*$ and consequently $\tau(\lambda) \equiv \tau^*$ where τ^* is the unique invariant measure of P^*. Therefore the condition to ensure convergence in Proposition 6.4 reduces to $2\overline{\eta}\alpha < 1$ (by considering the same Lyapunov function as in Proposition 4.5).

Notice that, in this case, $\nabla_x \overline{\Psi}_x(x_*, \lambda_*)$ is an average (by τ^*) of the Jacobian matrices $\nabla_x \overline{\Psi}(x_*, \lambda_*, w)$, where

$$\overline{\Psi}_x^{is}(x, \lambda, w) = \frac{\sigma^{is}(x^i)}{\lambda^{is}}\big[G^i(x, \sigma^{-i}(x), w) - x^{is}\big],$$

for every $w \in W$. Then the estimate given by (5.4) can be obtained exactly in the same manner than in Lemma 5.3 for the function $\overline{\Psi}$ if $2\overline{\eta}\alpha < 1$. Hence, as in the deterministic environment case, the matrix $\nabla\overline{\Psi}(x, \lambda)$ is stable for every rest point (x, λ) of (6.3) if $1 \leq 2\overline{\eta}\alpha < 2$ and Proposition 5.11 can also be recovered in this framework. Let $\overline{\mathcal{Y}}$ be the set of rest points of the dynamics (6.3).

Proposition 6.5. *If an attractor $\mathcal{A}$ for the dynamics (6.3) verifies that $B(\mathcal{A}) \cap \overline{\mathcal{Y}} \neq \emptyset$, then $\mathbb{P}(\mathcal{L}(x_n, \lambda_n) \subseteq \mathcal{A}) > 0$. In particular, under the Logit decision rule decision (5.1), if $1 \leq 2\overline{\eta}\alpha < 2$, then the set $\overline{\mathcal{Y}}$ reduces to one point (x_*, λ_*) and $\mathbb{P}((x_n, \lambda_n) \to (x_*, \lambda_*)) > 0$.*

Proof. We only have to check that the conclusion of Lemma 5.10 (regarding the attainability points of the process (APDM)) is satisfied in this case, since the remaining assumptions to apply [3, Theorem 7.3] hold just as before. For the sake of simplicity, let us suppose that τ^*_w is rational for all $w \in W$. Let (x, λ) be such that $x^{is} = G^i(s, \lambda^{-i}, \tau^*)$ for all $i \in A$ and $s \in S^i$. Take $m \in \mathbb{N}$ large such that, for every $w \in W$, $m\tau^*_w \in \mathbb{N}$. Then, for a fixed w, repeat the block-argument in the proof of Lemma 5.10 to construct a w-block of play. Then consider the block of play that consists on playing $m\tau^*_w$-times each w-block, for each w. This sequence generated by (APDM) carries the process $(x_n, \lambda_n)_n$ close to (x, λ). ☐

Remark 6.6. In fact, without the constantness assumption, we have shown that for every $(x, \lambda) \in X \times \Delta$ such that $x^{is} = G^i(s, \sigma^{-i}(x), \tau(\lambda))$ for all $i \in A$ and $s \in S^i$, the point (x, λ) is attainable by (APDM). Hence, in particular, any rest point of (6.3) is attainable and the first part in the statement of Proposition 6.5 holds only under condition (A) on the decision rule σ.

Acknowledgements

This article was written during the course of my PhD thesis. I would like to thank my advisor Sylvain Sorin for presenting this problem to me and also Michel Benaïm, Roberto Cominetti, and Mathieu Faure for very helpful discussions and comments.

References

[1] A. W. Beggs, *On the convergence of reinforcement learning*, J. Econom. Theory **122** (2005), 1–36.

[2] M. Benaïm, *Vertex-reinforced random walks and a conjecture of Pemantle*, Ann. Probab. **25** (1997), 361–392.

[3] ———, *Dynamics of stochastic approximation algorithms*, Séminaire de Probabilités, XXXIII, Lecture Notes in Math., vol. 1709, Springer, Berlin, 1999, pp. 1–68.

[4] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*, Springer-Verlag, Berlin, 1990.

[5] T. Börgers and R. Sarin, *Learning through reinforcement and replicator dynamics*, J. Econom. Theory **77** (1997), 1–14.

[6] O. Brandière and M. Duflo, *Les algorithmes stochastiques contournent-ils les pièges?*, Ann. Inst. H. Poincaré Probab. Statist. **32** (1996), 395–427.

[7] G. W. Brown, *Iterative solution of games by fictitious play*, Activity Analysis of Production and Allocation, John Wiley & Sons Inc., New York, N. Y., 1951, pp. 374–376.

[8] H. F. Chen, *Stochastic approximation and its applications*, Kluwer Academic Publishers, Dordrecht, 2002.

[9] R. Cominetti, E. Melo, and S. Sorin, *A payoff-based learning procedure and its application to traffic games*, Games Econ. Behav. **70** (2010), 71–83.

[10] Charles Conley, *Isolated invariant sets and the Morse index*, American Mathematical Society, Providence, R.I., 1978.

[11] M. Duflo, *Random iterative models*, Applications of Mathematics (New York), vol. 34, Springer-Verlag, Berlin, 1997.

[12] I. Erev and A. E. Roth, *Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria*, Amer. Econ. Rev. **88** (1998), 848–81.

[13] Y. Freund and R. E. Schapire, *Adaptive game playing using multiplicative weights*, Games Econom. Behav. **29** (1999), 79–103.

[14] D. Fudenberg and D. K. Levine, *The theory of learning in games*, MIT Press, Cambridge, MA, 1998.

[15] S. Hart and A. Mas-Colell, *A simple adaptive procedure leading to correlated equilibrium*, Econometrica **68** (2000), 1127–1150.

[16] ______, *A reinforcement procedure leading to correlated equilibrium*, Economics essays: A Festschrift for Werner Hildebrand, Springer, Berlin, 2001, pp. 181–200.

[17] H. J. Kushner and G. Yin, *Stochastic approximation and recursive algorithms and applications*, Springer-Verlag, New York, 2003.

[18] J. F. Laslier, R. Topol, and B. Walliser, *A behavioral learning process in games*, Games Econom. Behav. **37** (2001), 340–366.

[19] D. S. Leslie and E. J. Collins, *Individual Q-learning in normal form games*, SIAM J. Control Optim. **44** (2005), 495–514.

[20] J. W. Milnor, *Topology from the differentiable viewpoint*, Princeton University Press, Princeton, NJ, 1997.

[21] R. Pemantle, *Nonconvergence to unstable points in urn models and stochastic approximations*, Ann. Probab. **18** (1990), 698–712.

[22] M. Posch, *Cycling in a stochastic learning algorithm for normal form games*, J. Evol. Econ. **7** (1997), 193–207.

[23] H. Robbins and S. Monro, *A stochastic approximation method*, Ann. Math. Statistics **22** (1951), 400–407.

[24] S. J. Schreiber, *Urn models, replicator processes, and random genetic drift*, SIAM J. Appl. Math. **61** (2001), 2148–2167.

[25] A. N. Shiryaev, *Probability*, Springer-Verlag, New York, 1996.